Molecular cloning of a cDNA encoding alliinase from onion (*Allium cepa* L.)

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Molecular Biology

in the

University of Canterbury

by

Sonya A. Clark

1993

University of Canterbury,

Christchurch, New Zealand.

Life itself is like an onion; it has a bewildering number of layers.
You peel them off, one by one, and sometimes you cry.

Carl Sanburg.

Molecular cloning of a cDNA encoding alliinase from onion (*Allium cepa* L.)

Abstract

Alliinase catalyses the release of a range of volatile sulphur compounds which are responsible for the distinctive flavour and odour of *Allium* species. Clones encoding this vacuolar enzyme were isolated by immunoscreening a cDNA expression library constructed in the vector λZAPII with mRNA extracted from sprouting *A. cepa* bulb shoots. Six clones were characterized by DNA sequencing. Four of the cDNA sequences (Alli6, 7, 8B and 9) were found to be identical apart from variation in the length of their 5' ends. These sequences contained three putative polyadenylation signals. A fifth clone, Alli4B, was also very similar but displayed a truncated 3' end with six divergent bases just prior to the polyadenylate tail and lacked the most 3' polyadenylation signal present within the sequences of the other clones. It was evident from these differences that the clones were unique. However, their high sequence homology suggested they were encoded by a single gene, or two that are very closely related. The sixth clone sequenced, Alli4A, showed no homology to the other five clones, nor to any sequences within the GenBank database. The five highly homologous cDNA sequences ranged in length from 1604 bp to 1757 bp, and all appeared to contain a complete longest open reading frame encoding a polypeptide with a predicted size of 54 884 Da.

Four peptide sequences derived from purified *A. cepa* alliinase were aligned, and showed 93% homology with the corresponding amino acid sequences deduced from the cDNA clones. Alignment of the native alliinase N-terminal peptide to the cDNA-inferred protein sequence predicted a hydrophobic 34-residue prepeptide sequence terminating in a peptidase cleavage site. Given the vacuolar location of this enzyme, it is probable that this region functions in targeting of the alliinase precursor to the endoplasmic reticulum. The inferred mature alliinase subunit polypeptide contained sequence motifs compatible with both Asn-linked glycosylation and pyridoxal phosphate cofactor binding.

Conversion of clones to the phagemid form allowed the protein expressed in *Escherichia coli* from Alli6 to be western blotted. This analysis revealed protein moieties with molecular masses of 47 and 41.4 kDa. Southern hybridization analysis of *A. cepa* genomic DNA using an alliinase cDNA probe demonstrated the presence of a small multigene family with at least four members. Further genomic DNA hybridization analysis using a probe encompassing the 3' untranslated region of the

alliinase clones demonstrated that the highest identity was to a single fragment, suggesting some alliinase gene family members may be transcriptionally inactive in sprouting bulb shoot tissue.

A 1.7 kb transcript was detected by northern analysis of RNA extracted from developing *A. cepa* seedlings, indicating that the cDNA clones were near-full-length. Alliinase mRNA could be easily detected in seedlings up to 14 days after germination. Visual assessment and densitometer analysis of hybridization intensities suggested that alliinase transcripts were most abundant two to six days after germination and declined rapidly over the ensuing eight days. This indicated that alliinase expression is regulated at the transcriptional level during this period.

Table of contents **Page**

Chapter 4. Discussion

Chapter 5. Concluding summary and discussion

List of figures

List of tables

Abbreviations used in this thesis

ATP	adenosine triphosphate
BSA	bovine serum albumin
bp	base pairs
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
CAT	chloramphenicol acetyltransferase
CSO	cysteine sulphoxide
CTAB	cetyltrimethylammonium bromide
C-terminus	carboxyl terminus
Da	daltons
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DEPC	diethylpyrocarbonate
dGTP	deoxyguanosine triphosphate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTE	dithioerythritol
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
GAR-AP	goat anti-rabbit alkaline phosphatase
GPNA	γ-glutamyl-ρ-nitroanilide
GUS	β-glucuronidase
HCC	hexamine cobalt chloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
IOD	integrated optical density
IPTG	isopropyl-1-thio-β-D-galactoside
kb	kilobases
kbp	kilobase pairs
kDa	kilodaltons
K_m	Michaelis constant
LDH	lactate dehydrogenase
MES	4-morpholineethanesulphonic acid
MOPS	4-morpholinepropanesulphonic acid
mRNA	messenger RNA
NADH	nicotinamide adenine dinucleotide
NBT	4-nitro blue tetrazolium chloride
N-terminus	amino terminus
oligonucleotide	oligodeoxyribonucleotide
ORF	open reading frame
PCR	polymerase chain reaction
PEG	polyethylene glycol

pfu	plaque forming units
PLP	pyridoxal phosphate
rDNA	ribosomal DNA
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
TCA	trichloroacetic acid
TEMED	N,N,N',N'-tetramethylethylenediamine
TFMS	trifluoromethanesulphonic acid
TLC	thin layer chromatography
TMAC	tetramethylammonium chloride
URF	upstream reading frame
UTR	untranslated region
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactoside

Chapter 1. Introduction

1.1. Overview

Allium cepa (onion) is one of the oldest known cultivated vegetables, having been grown by man for more than 4000 years (Fenwick & Hanley 1985). *A. cepa*, together with its close relative *Allium sativum* (garlic), has been used extensively to enhance food flavour and as a home remedy for a variety of illnesses. References to both of these plants can be found in medicinal literature dating back to Hippocrates in the 5th century B.C. (Hanelt 1990). This long history of culinary and medicinal use prompted researchers in the 1950s to attempt the isolation and identification of the endogenous substances responsible for the antibiotic effects exhibited by *Allium* extracts (eg. Virtanen & Matikkala 1959). This work in turn led to a focus on the sulphur compounds of *Allium* species, many of which have a role in flavour production. In 1947, Stoll & Seebeck established that *A. sativum* bulbs contained large amounts of a unique sulphur-containing amino acid which they termed 'alliin' (*S*-(2-propenyl)-L-cysteine sulphoxide, or *S*-allyl-cysteine sulphoxide). In 1951 these same authors further demonstrated that *A. sativum* bulbs contained an enzyme which they named alliinase (E.C. 4.4.1.4), that converts alliin to 'allicin' (later determined to be diallyl thiosulphinate). This reaction is central to the flavour biochemistry of *Allium* species. Since this time the biosynthesis and chemistry of the sulphur compounds of the *Allium* species have been extensively investigated.

Current studies of many plant biosynthetic pathways are directed toward identification and characterization of the genes involved and analysis of their expression. The expression of a gene product can be affected at many levels, and examples of transcriptional, post-transcriptional, translational and post-translational control can be found. The control of plant gene expression can only be discussed with reference to gene organization and structure, and to hypotheses related to the control of eukaryotic gene transcription, the role of *cis*- and *trans*-acting factors and the effects of developmental and environmental cues. In this introduction, the flavour biochemistry in *Allium* species with particular reference to the role of alliinase will be reviewed. The structure of plant nuclear genes will then be described and the relationship between gene structure and function discussed.

1.2. *Allium cepa* flavour biochemistry and the role of alliinase

1.2.1. Economic importance of *Allium* species

Edible *Allium* species are of major economic and dietary importance throughout the world. A range of *Allium* species have significant economic value in specific areas but onion and garlic are grown, traded and consumed in most countries. As a result, selection and breeding of locally adapted *A. cepa* and *A. sativum* cultivars has led to their cultivation from the tropics to subarctic regions. A 1990 estimate placed the value of bulb onion production at $US 5 billion annually, and there is a large world trade of which New Zealand is a small part. In 1990 approximately 69,800 tonnes of fresh onions with an export value of $NZ 37.5 million were exported from New Zealand, with almost one third of these going to Japan (McPherson *et al.* 1992a). Two New Zealand-bred cultivars with good storage characteristics, 'Pukekohe Longkeeper' and 'Early Longkeeper' are most commonly exported. However, both of these cultivars have been criticised for their high pungency and lack of sweetness, particularly by Japanese consumers (McPherson *et al.* 1992a). As onions are mainly grown, processed and consumed for their distinctive taste and aroma, production of bulbs with a flavour preferred by consumers may provide wider markets for New Zealand onion growers. Plant breeding approaches have been highly successful in developing varieties such as 'Pukekohe Longkeeper' with long-lasting characteristics. However, progress in the development of a milder flavoured 'PLK' has been slow.

1.2.2. *A. cepa* - taxonomy and genetic structure

Allium and related genera have been placed in several families by different taxonomists. In early classifications of the angiosperms *Allium* were placed in the family *Liliaceae* (Cronquist 1968). However, some taxonomists have included them in the *Amaryllidaceae* (Hutchinson 1973). In a more recent taxonomic treatment of the monocotyledons, *Allium* and closely related genera are placed into the *Alliaceae* family (Dahlgren *et al.* 1985), which is close to the family *Amarylidaceae*.

Therefore the adopted hierarchy (Dahlgren *et al.* 1985) is:

1. Class	*Monocotyledones*
2. Superorder	*Liliflorae*
3. Order	*Asparagales*
4. Family	*Alliaceae*
5. Tribe	*Allieae*
6. Genus	*Allium*

Although the domesticated edible forms number only seven main species, more than 600 species are estimated to be present in the *Allium* genus (Hanelt 1990). *A. cepa* is the type species of the subgenus *Allium*, and is diploid, with the basic chromosome number $x = 8$ ($2n = 16$). The 2C nucleus contains 32.2 pg of DNA, equivalent to 3.1×10^{10} bp (Arumuganathan & Earle 1991), with a G+C content of 34.6% (Kirk *et al.* 1970). Up to 50% of this DNA is in the form of repetitive sequences (Stack & Comings 1978). This haploid genome size is large by comparison to other flowering plants. For comparison, the equivalent 2C value in *Triticum aestivum* is 3.2×10^{10} bp and in *Phaseolus vulgaris* is 1.3×10^{10} bp (Arumuganathan & Earle 1991).

1.2.3. *Allium* flavour biochemistry

1.2.3.1. Flavour precursors of the *Allium* species

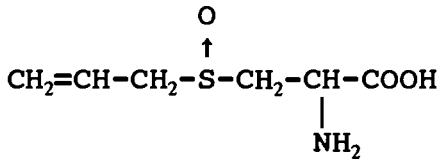
Interest in the compounds responsible for the flavour of *Allium* species led to the isolation of some sulphide compounds from garlic and onion oil by the middle of the nineteenth century. The identification of flavour precursors and their isolation from intact *A. cepa* and *A. sativum* tissues eventually came during the period 1955-1970 through the work of Virtanen, Suzuki and co-workers.

A notable feature of the chemical composition of *Allium* plants is the abundance of organically bound sulphur, which is considerably greater than that found in most organisms (Whitaker 1976). A large proportion of this is found incorporated into unique nonprotein sulphur amino acids, collectively referred to as the *S*-L-alk(en)yl cysteine sulphoxides (*S*-alk(en)yl-CSOs). The characteristic flavour of *Allium* species results from the catalytic hydrolysis of *S*-alk(en)yl-CSOs by the enzyme alliinase, resulting in the release of a wide range of volatile sulphide compounds. This reaction is central to the flavour biochemistry of *Allium* species and over 80

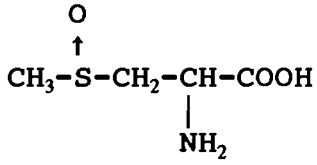
volatile compounds which contribute to flavour have been reported in fresh, cut, and steam-distilled extracts of *A. cepa* and *A. sativum* (Boelens *et al.* 1971, Freeman 1975). However it is the identity and relative proportions of the various *S*-alk(en)yl-CSO precursors within the cell that ultimately determines flavour potential.

Four *S*-alk(en)yl-CSOs are most commonly found in *Allium* species. These are;

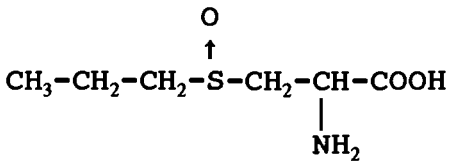
1. (+)-*S*-(2-propenyl)-L-cysteine sulphoxide (*S*-allyl-CSO, also known as alliin).



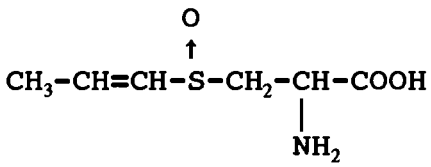
2. (+)-*S*-methyl-L-cysteine sulphoxide (*S*-methyl-CSO)



3. (+)-*S*-propyl-L-cysteine sulphoxide (*S*-propyl-CSO)



4. *trans*-(+)-*S*-(1-propenyl)-L-cysteine sulphoxide (*S*-1-propenyl-CSO)



These compounds contain a potentially diasteromeric sulphoxide bond, but the naturally occurring compounds are all (+) isomers (Whitaker 1976).

1.2.3.1.1. Flavour precursor distribution in plants

The first of the *S*-alk(en)yl-CSOs isolated, *S*-allyl-CSO, was extracted from *A. sativum* and *A. ursinum* tissues (Stoll & Seebeck 1947). It occurs in amounts up to 2.4 g/kg in *A. sativum* (Stoll & Seebeck 1951) but at lower levels in most of the other *Allium* species in which it has been found (Freeman & Whenham 1975). *S*-Methyl-CSO and *S*-propyl-CSO were first isolated from *A. cepa* (Virtanen & Matikkala 1959). *S*-Methyl-CSO is the predominant flavour precursor in many ornamental *Allium* species and has a broad distribution, being found in all *Allium* species examined. *S*-Propyl-CSO is also widely distributed among *Allium* species as well as being reported in *Ipheion uniflorum* (Fujiwara *et al.* 1958) and tends to be present at low concentrations. *S*-1-Propenyl-CSO was isolated from *A. cepa* in 1961 (Virtanen and Spåre 1961) during an investigation into the lachrymatory (tear-producing) properties of onions and is found in many *Allium* species. Various sulphoxide and thiolether derivatives of the *S*-alk(en)yl-substituted cysteines are also found in other members of the *Alliaceae*, and in the families *Liliaceae* and *Crucifereae*. They have also been reported in some members of the families *Compositae*, *Umbelliferae* and *Leguminosae* (Fowden 1964, Whitaker 1976).

1.2.3.1.2. Flavour precursor composition in *Allium* species

The *Allium* species analysed to date can be classified into distinct groups based on their flavour precursor composition. It has been proposed that this could be used as an additional characteristic in taxonomic classifications (Lancaster & Boland 1990). Several classifications have been suggested by different authors (eg. Bernhard 1970, Freeman & Whenham 1975). Lancaster & Boland (1990) have proposed four classes based on the presence or absence of *S*-1-propenyl- and *S*-allyl-CSOs. Group 1 is characterized by the presence of *S*-methyl-, *S*-1-propyl- and *S*-1-propenyl-CSOs, and the absence of *S*-allyl-CSO. This group includes *A. cepa*, leek (*Allium porrum*) and many ornamental *Allium* species. The predominant flavour precursor of *A. cepa* is *S*-1-propenyl-CSO (Schwimmer 1969). *S*-Propyl-CSO and *S*-methyl-CSO are also present at lower concentrations. *Allium* species in group 2 are characterized by the presence of *S*-methyl-, *S*-1-propyl- and *S*-allyl-CSOs and the absence of *S*-1-propenyl-CSO. Included in this group is *A. sativum*, in which *S*-allyl-CSO is the predominant flavour precursor. *Allium* species which contain neither *S*-1-propenyl-CSO nor *S*-allyl-CSO are classified into group 3, and those containing all four flavour precursors are classified into group 4. Only *Allium*

species which produce *S*-1-propenyl-CSO have a lachrymatory effect, while *Allium* species containing predominantly *S*-allyl-CSO tend to have a garlic-like flavour and odour.

1.2.3.1.3. Cultivar variation and environmental and developmental effects on *A. cepa* flavour precursor abundance

The flavour of an *A.cepa* bulb is primarily determined by its genotype, and ten-fold differences in flavour levels have been found between various *A. cepa* cultivars (Lancaster & Boland 1990). *A. cepa* cultivars which have been bred for high dry matter and processing (e.g. 'Dehyso') have some of the highest flavour levels, whereas some of the Japanese cultivars are very mild (Lancaster & Boland 1990). Temperature, water supply, and sulphur nutrition all affect the flavour strength of bulbs. Onions grown at temperatures between 21°C and 27°C were found to contain three times the amount of volatile sulphur compounds of those grown between 10°C and 15°C (Platenius 1944). This may have been due in part to faster growth and development of those plants at warmer temperatures. Bulbs tend to be more pungent when soil water is low, and increasing the level of sulphur in the soil also increases their flavour levels (Paterson 1979, Freeman 1979).

The proportions of the various *S*-alk(en)yl-CSOs in *A. cepa* bulbs have been shown to vary with developmental stage, and with cultural and climatic conditions. *A. cepa* seeds have low levels of flavour precursors but significant amounts of γ-glutamyl peptides which are believed to be metabolized to free flavour precursors during germination (Matikkala & Virtanen 1965b, McCallion & Lancaster 1984). Flavour precursors accumulate rapidly throughout germination and emergence of *A. cepa* cotyledons and are particularly high in the hypocotyl (Freeman 1979). The relative proportions of the various *S*-alk(en)yl-CSOs have also been found to differ in *A. cepa* seedlings grown with or without light. Under normal light conditions the dominant flavour precursor, *S*-propyl-CSO, was detected at highest levels in the shoots of seedlings. However *S*-1-propenyl-CSO was dominant during very early seedling growth with high concentrations detected in the root and hypocotyl. When *A. cepa* seedlings were grown in the dark the ratios of the *S*-alk(en)yl-CSOs changed and *S*-methyl-CSO, which is a minor component of the flavour precursors present in light-grown seedlings, was found to represent a greater proportion of the flavour precursor pool during early cotyledon development (McCallion & Lancaster 1984).

Flavour precursor accumulation is thought to be initially reliant on resources from the seed, but appears to become dependent on photosynthetic products as the

cotyledonary leaf develops (McCallion & Lancaster 1984). Younger leaf blades were found to produce more flavour precursors than older leaf blades, and it is believed that leaf blades are the site of synthesis of flavour precursors which are then transported to, and stored in, the scales of the bulb (Lancaster *et al.* 1986). Flavour precursor content increases in scales during bulbing and then gradually decreases as the bulbs mature (Lancaster *et al.* 1984).

In vitro cell cultures of differentiated *A. cepa* tissue were shown to produce flavour precursors similar to those of the original bulb explants, but undifferentiated callus tissues contained only small amounts of flavour components (Freeman *et al.* 1974, Selby *et al.* 1979, Turnbull *et al.* 1981). *A. sativum* tissue has also been reported to change flavour when grown in culture. Only *S*-methyl-CSO was detected by Lancaster *et al.* (1988a) in undifferentiated white callus, whereas green callus tissues contained all three of the flavour precursors normally found in differentiated *A. sativum* tissues. This observation led to a proposal that expression of *S*-allyl- and *S*-propyl-CSOs requires the presence of plastids, possibly chloroplasts, whereas *S*-methyl-CSO appeared to be regulated differently.

1.2.3.1.4. Location of flavour precursors in plant tissues

S-Alk(en)yl-CSOs have been reported to be present in the bulbs, leaf blades, base plates and roots of *A. cepa*, *A. porrum* and *A. sativum* (Freeman 1975, Lancaster *et al.* 1986), but are absent from the seeds (Freeman 1979). The scape and inflorescence of *A. cepa* and *A. sativum* have been found to produce the characteristic sulphide flavour when eaten, indicating the presence of both alliinase and flavour precursors. However, flavour is negligible in the cortical pith of the leaf blades and the flowering scape (Lancaster & Boland 1990).

The concentration of flavour precursors varies within *A. cepa* bulb scales and leaf blades; younger tissues generally show high levels, and progressively decreasing levels have been detected in older tissues (Freeman 1975, Lancaster *et al.* 1986). A similar distribution of flavour precursors was also found in *A. porrum* tissues (Freeman 1975). Studies based on histochemical localization of sulphur compounds in *A. cepa* bulb scales have shown that the bundle sheath contains a higher concentration of flavour precursor peptides than the surrounding cortical cells. This was thought to indicate transport or synthesis of flavour precursors within the bundle sheath (Becker & Schuphan 1975). Flavour precursor formation in *A. cepa* and *A. sativum* cell cultures has been correlated with the presence of vesicles in the cytoplasm, and it has been suggested that these may contain the flavour precursors (Turnbull *et al.* 1981, Mohanty *et al.* 1989).

1.2.3.1.5. Biosynthesis of flavour precursors

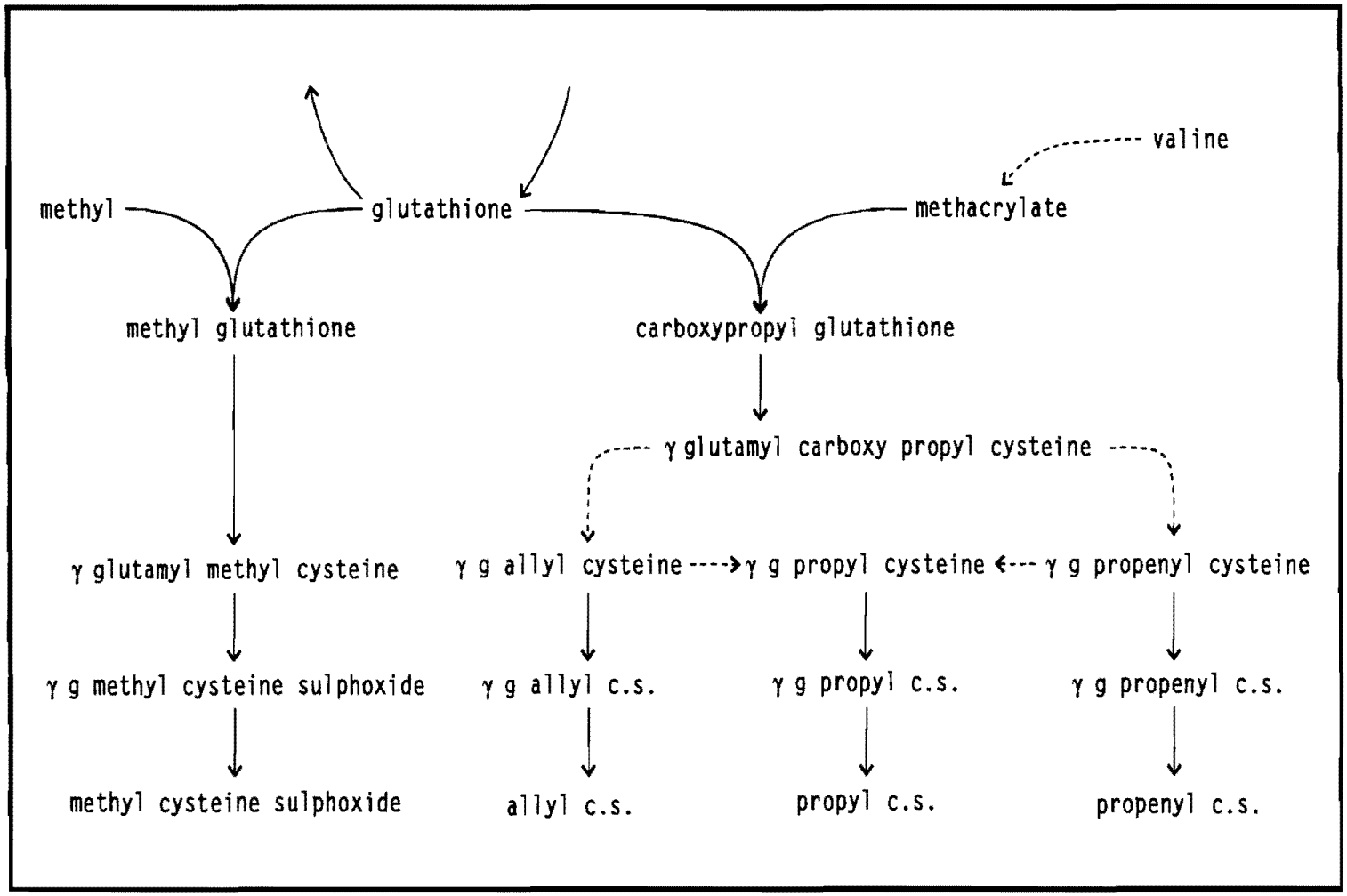
The biosynthesis of the *S*-alk(en)yl-CSO flavour precursors and other nonprotein sulphur amino acids involves the interaction of the carbon, nitrogen and sulphur pathways. Sulphur is taken up from the soil in the form of sulphate (SO_4^{2-}), and is transported through the xylem from the roots to leaves where it is reduced to sulphide and assimilated into cysteine. Some sulphate reduction and assimilation can also occur in the roots. Cysteine is the principal starting metabolite for the synthesis of sulphur-containing compounds and is converted to protein cysteine, protein methionine, methionine and glutathione via various intermediates. In most plants 90% of the sulphur present is in the form of cysteine and methionine, which are themselves incorporated into proteins (for a review of sulphur reduction and assimilation in plants see Giovanelli *et al.* 1980). In *Allium* species the majority of the sulphur is present in the form of nonprotein amino acid derivatives (Fenwick & Hanley 1990).

The early stages of the *A. sativum* flavour precursor biosynthetic pathways were elucidated by studies initiated in the 1960s by Suzuki and colleagues utilizing radiolabelled sulphur compounds (Suzuki *et al.* 1962, Sugii *et al.* 1963a, 1963b). These investigations showed that uptake of labelled sulphur resulted in many labelled γ-glutamyl peptides, as well as flavour precursors. Support for these results came from parallel studies by Virtanen, Granroth and colleagues on the biosynthesis of flavour precursors in *A. cepa* (Ettala & Virtanen 1962, Virtanen 1965, Granroth 1970). Further investigations into the biosynthesis of flavour precursors in *A. cepa* by Lancaster *et al.* (1988b) showed that ^{35}S-labelled sulphate was ultimately incorporated into free flavour precursors, and suggested that the γ-glutamyl peptides are intermediates in the pathway. This work led to the proposed flavour precursor pathway shown in Figure 1.1 (Lancaster & Shaw 1989). All the intermediates in this pathway, with the exception of γ-glutamyl *S*-allyl-CSO, have been reported in *A. cepa*, *A. sativum* or *A. schoenoprasum* (chive) (Lancaster & Boland 1990). It is believed that the biosynthesis of glutathione occurs in the chloroplasts, and the subsequent metabolism from glutathione via the γ-glutamyl peptides to flavour precursors is cytoplasmic (Lancaster *et al.* 1988b). The biosynthesis and complex chemistry of the sulphur compounds of *Allium* species has been recently reviewed by Block (1992).

Figure 1.1. Proposed biosynthetic pathway of the flavour precursors of *Allium* species

γ g = γ-glutamyl, c.s. = cysteine sulphoxide.

Reproduced with permission from Lancaster & Boland (1990).


valine
methyl
glutathione
methacrylate
methyl glutathione
carboxypropyl glutathione
γ glutamyl carboxy propyl cysteine
γ glutamyl methyl cysteine
γ g allyl cysteine
γ g propyl cysteine
γ g propenyl cysteine
γ g methyl cysteine sulphoxide
γ g allyl c.s.
γ g propyl c.s.
γ g propenyl c.s.
methyl cysteine sulphoxide
allyl c.s.
propyl c.s.
propenyl c.s.

1.2.3.1.6. Role of the flavour precursors of *Allium* species

The function of many plant secondary products is unknown, but they are generally believed to function as storage, transport or chemical defence compounds (phytoalexins). The nonprotein amino acids are an integrated part of the metabolic network of sulphur compounds in *Allium* species. γ-Glutamyl peptides have a nitrogen:sulphur ratio of 2:1, are found in high levels in storage organs such as dormant bulbs and seeds, and are rapidly metabolized at germination (Matikkala & Virtanen 1965a, 1965b, Schwimmer & Austin 1971, McCallion & Lancaster 1984, Fenwick & Hanley 1985). It has been hypothesized that these peptides may have a role in the storage of nitrogen and sulphur, which can be mobilized rapidly for use in the biosynthesis of essential compounds during bulb sprouting or seed germination (Freeman & Whenham 1976a, 1976b). Kasai & Larson (1980) have also postulated that γ-glutamyl derivatives may function in the transport of amino acids across membranes. The flavour precursors of *Allium* species, in conjunction with alliinase, have been postulated to function in defence (Mazelis & Fowden 1973). This hypothesis is discused in Subsection 1.2.3.2.2.g.

1.2.3.2. Enzymes involved in *A. cepa* flavour production

1.2.3.2.1. γ-Glutamyl transpeptidase/peptidase

Whilst numerous enzymes are involved in the biosynthesis of flavour precursors in *Allium* species, only γ-glutamyl transpeptidase/peptidase (E.C. 2.3.2.2) has been investigated in detail. Among the *Allium* species, 24 γ-glutamyl peptides have been found, 18 of which contain sulphur (Lancaster & Boland 1990). *A. cepa* and *A. sativum* both contain significant proportions of γ-glutamyl peptides. The penultimate enzyme in the proposed flavour release pathway, γ-glutamyl transpeptidase/peptidase (GTP/GP) can catalyse the hydrolysis of γ-glutamyl linkages and the transfer of the γ-glutamyl group to a large number of amino acid and peptide acceptors. GP activity does not require γ-glutamyl acceptors and so is a true hydrolase. Both of these activities are capable of hydrolysing the γ-glutamyl-CSOs of *Allium* species to produce the *S*-alk(en)yl-CSOs which are then available as substrates for alliinase (see Figure 1.1). Whether these activities are attributable to the presence of two enzymes, or to a single enzyme with a dual function has not been resolved.

Neither GTP or GP activities are detectable in dormant *A. cepa* bulbs, but both are found in bulbs that have sprouted (Austin & Schwimmer 1971, Schwimmer & Austin 1971). Accumulation of γ-glutamyl peptides in stored *A. cepa* bulbs (Matikkala & Virtanen 1965b) is consistent with lack of GTP/GP expression.

During sprouting, GTP activity has been detected at high levels (Austin & Schwimmer 1971, Schwimmer & Austin 1971), which may serve to metabolize these peptides to free flavour precursors (Lancaster & Boland 1990).

An enzyme which showed GTP activity was purified 800-fold from sprouting *A. cepa* bulbs (Schwimmer & Austin 1971). This enzyme was identified as a transpeptidase because of the enhancement in the rate of release of ρ-nitroaniline (GPN) from the synthetic substrate γ-glutamyl-ρ-nitroanilide (GPNA) by amino acids, and the competitive inhibition of the enzyme by γ-glutamyl derivatives. It was also found to have a pH optimum of 9.0, and to be inhibited by borate. GP which was purified about 60-fold from germinating *A. schoenoprasum* seeds, displayed a pH optimum of 8.0 and was also inhibited by borate (Matikkala & Virtanen 1965a). Although no GTP activity could be detected in this preparation, GTP and GP activities were detected in aqueous extracts of sprouted *A. sativum* cloves (Ceci *et al.* 1992).

When GPNA is used as the substrate, extracts of kidney bean (*Phaseolus vulgaris*) and ackee plant (*Blighia sapida*) display a transferase activity at higher pH (9.5) and a hydrolase activity at lower pH (6.5) (Goore & Thompson 1967, Kean & Hare 1980). Ceci *et al.* (1992) postulated that the bimodal rate of cleavage of GPNA shown by the GTP/GP activity in an aqueous extract of *A. sativum* indicated that two enzymes with separate functions are present in *Allium* species. However, enzyme preparations from *A. cepa* which have been purified to apparent homogeneity display both activities (M.L. Shaw pers. comm.). Hence, it is thought that the hydrolase activity may function in the biosynthesis of flavour precursors, whereas the transpeptidase activity is only induced under conditions of unphysiologically high pH *in vitro* (J.E. Lancaster pers. comm.).

1.2.3.2.2. Alliinase

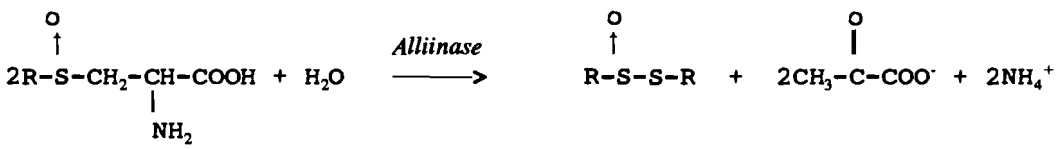
Alliinase is the enzyme ultimately responsible for the development of the volatile sulphur-containing flavour compounds of *Allium* species, and both the *A. cepa* and *A. sativum* alliinases have been extensively investigated. Alliinase was first identified by Stoll & Seebeck in 1951 when they demonstrated an enzyme activity which hydrolysed *S*-allyl-CSO in an aqueous extract of *A. sativum*. This enzyme is classified as alliin lyase (EC 4.4.1.4), and is also known as alliin alk(en)yl sulphenate lyase or alliinase, or by the synonyms *S*-alk(en)yl-L-cysteine sulphoxide lyase, cysteine sulphoxide lyase, or C-S lyase.

1.2.3.2.2.a. Activity

The reaction catalysed by alliinase is a β-elimination of the *S*-alk(en)yl sulphoxide group from the *S*-alk(en)yl-CSO substrate:

S-alk(en)yl-CSO | *S*-alk(en)yl sulphenic acid | α-iminopropionic acid

Both the products of this reaction are chemically unstable. The ketimine product spontaneously hydrolyses to pyruvate and ammonia, whilst the reactive sulphur species can combine with a range of coreactants, most frequently another of the same species, to result in a wide range of thiosulphinate derivatives. Therefore this reaction is commonly described as:



S-alk(en)yl-CSO | dialk(en)yl thiosulphinate | pyruvate | ammonia

1.2.3.2.2.b. Flavour release by alliinase

Flavour release in *A. sativum* involves the action of alliinase on *S*-allyl-, *S*-methyl- and *S*-propyl-CSOs to initially produce the corresponding unstable sulphenic acid intermediates. The main flavour precursor of *A. sativum*, *S*-allyl-CSO, is cleaved by alliinase to produce allyl sulphenic acid which can then react non-enzymatically with another molecule of allyl sulphenic acid to form diallyl thiosulphinate (allicin). Allicin is the predominant thiosulphinate produced, and it undergoes nonenzymic rearrangements to form mono-, di-, trisulphides, thiosulphonates and sulphur dioxide (Brodnitz *et al.* 1971).

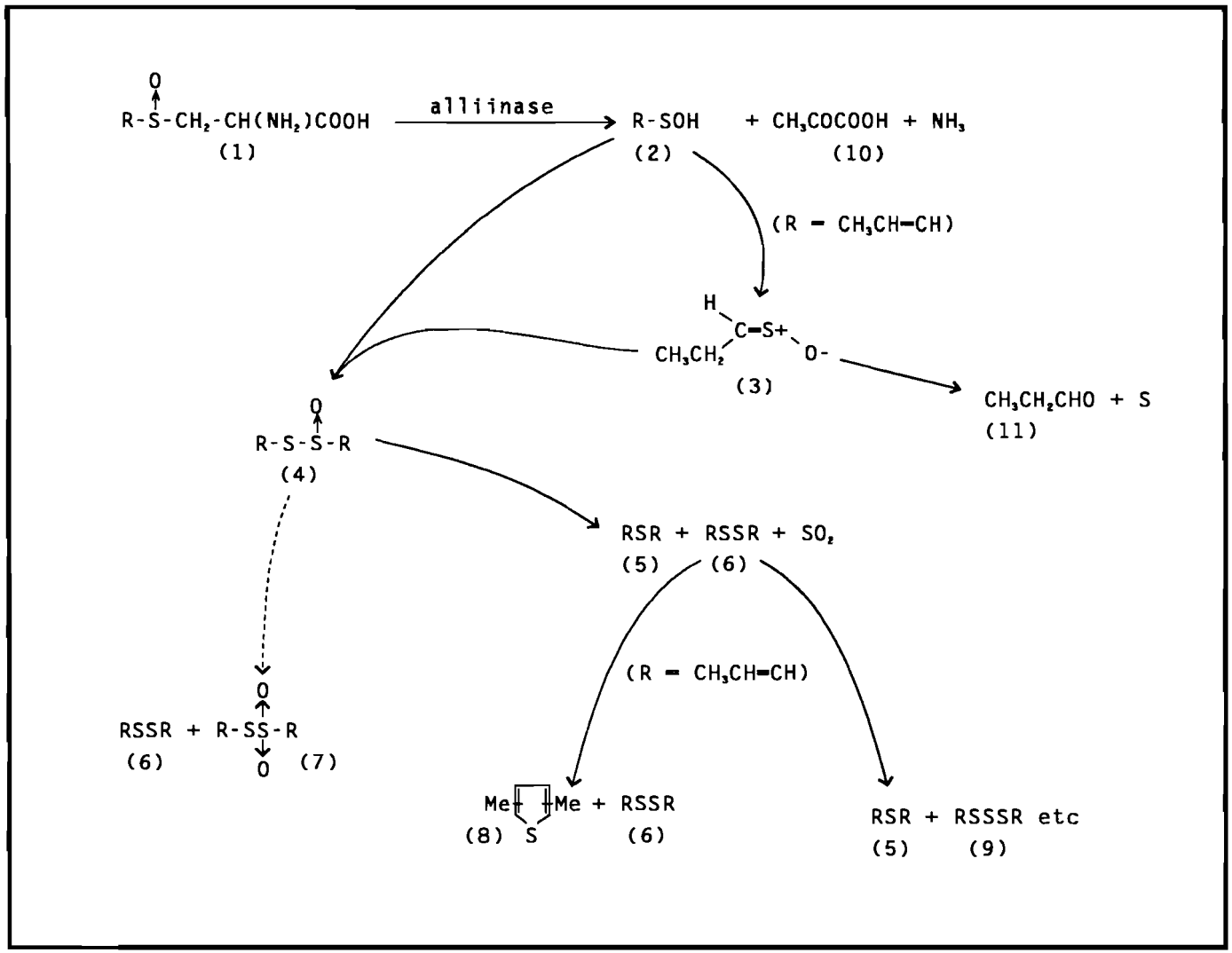
In *A. cepa* the predominant flavour precursor, *S*-1-propenyl-CSO, is hydrolysed by alliinase to form 1-propenyl sulphenic acid. Unlike the other alk(en)yl sulphenic acids which immediately recombine to form various thiosulphinates, 1-propenyl sulphenic acid initially rearranges to produce thiopropanal *S*-oxide which can then rearrange to form propanal, carbonyl derivatives of propanal, and elemental sulphur (Boelens *et al.* 1971, Brodnitz & Pascale 1971, Freeman & Whenham 1976a, 1976b). Thiopropanol *S*-oxide can also rearrange to form thiosulphinates (Figure 1.2).

Figure 1.2. Volatile sulphur compounds formed in *A. cepa* following cleavage of alk(en)yl-cysteine sulphoxides by alliinase

(1) *S*-alk(en)yl cysteine sulphoxide
(2) *S*-alk(en)yl sulphenic acid
(3) thiopropanal *S*-oxide
(4) thiosulphinate
(5) monosulphide
(6) disulphide
(7) thiosulphonate
(8) dimethyl thiophene
(9) trisulphide
(10) pyruvate
(11) propanal

R = methyl [CH_3], propyl [C_3H_7] or propenyl [CH_3-CH=CH].

Reproduced with permission from Lancaster & Boland (1990).


O
R-S-CH2-CH(NH2)COOH
(1)
alliinase
R-SOH
(2)
+ CH3COCOOH + NH3
(10)
(R = CH3CH=CH)
H
C=S+
CH3CH2
O-
(3)
CH3CH2CHO + S
(11)
O
R-S-S-R
(4)
RSR + RSSR + SO2
(5) (6)
(R = CH3CH=CH)
RSSR + R-SS-R
(6) O (7)
O
Me Me + RSSR
(8) S (6)
RSR + RSSSR etc
(5) (9)

Further conversions have also been observed, as are shown for *A. cepa* in Figure 1.2. The compound responsible for the lachrymatory effect of *A. cepa* was originally identified as 1-propenyl sulphenic acid (Virtanen & Spåre 1961). However, Wilkins proposed that the lachrymator was thiopropanal *S*-oxide (Wilkins 1961), and this structure was later confirmed by Brodnitz & Pascale (1971). As *A. sativum* does not contain *S*-1-propenyl-CSO, no lachrymatory effect is experienced when garlic tissues are crushed.

1.2.3.2.2.c. Alliinase and related enzymes

Three types of plant enzymes which cleave *S*-alk(en)yl-CSOs have been examined. The first type act specifically on *S*-alk(en)yl-CSOs, such as the *A. cepa* and *A. sativum* enzymes (referred to as 'alliinase' in this thesis). Alliinase enzymes are probably present in most if not all of the *Allium* species (Lancaster & Boland 1990) and have been highly purified from *A. sativum* (Kazaryan & Goryachenkova 1978), *A. cepa* (Tobkin & Mazelis 1979) and to various extents from *A. chinense* (rakkyo) (Tsuno 1958), *A. porrum* (Nock & Mazelis 1989) and *A. fistulosum* (welsh onion) (Fujita *et al.* 1990). A similar substrate specificity has been reported for an alliinase detected in aqueous extracts of *Tulbaghia violacea*, a plant which also releases odorous compounds when cells are damaged (Jacobsen *et al.* 1968). The thiolether analogues of the cysteine sulphoxide substrates competitively inhibit the activity of the *A. cepa* and *A. sativum* enzymes (Whitaker 1976).

The second group are classified as alk(en)yl-L-cysteine lyase (E.C. 4.4.1.6) and hydrolyse both sulphoxide and thiolether derivatives of *S*-alk(en)yl-substituted cysteines. Enzymes with this broader substrate specificity have been identified in the plants *Acacia farnesiana* (Mazelis & Creveling 1975, Sweet & Mazelis 1987), *Acacia georginae*, *Albizzia julibrissin* (Mazelis & Fowden 1973) and *Albizzia lophanta* (Schwimmer & Kjær 1960). The natural substrate of the *Acacia farnesiana* enzyme is thought to be L-djenkolate, an analogue of *S*-methyl-L-cysteine (Mazelis & Creveling 1975). The enzymes of *A. georginae* and *A. julibrissin* were found to show higher affinity for their various endogenous substrates than to substrate analogues from other species (Mazelis & Fowden 1973). An alk(en)yl-L-cysteine lyase has also been purified from the fruiting bodies of the shiitake mushroom, *Lentinus edodes*, in which the predominant substrate is desglutamyl-lentinic acid, produced by the hydrolysis of the γ-glutamyl moiety of lentinic acid (Iwami & Yasumoto 1980). An enzyme that shows an analogous activity has been identified in the bacterial species *Pseudomonas cruciviae* (Nomura *et al.* 1963) and *Bacillus*

subtilus (Murakami 1960), and purified from *Pseudomonas putida* (Kamitani *et al.* 1990). The natural substrate of the *P. putida* enzyme is also L-djenkolate (Kamitani *et al.* 1991).

A third related enzyme, cystine lyase, is found in a number of *Brassica* species (Mazelis 1963) and can cleave both *S*-alk(en)yl-CSOs and L-cystine. This enzyme has been partially purified and studied in the *B. oleracea* varieties cauliflower (De Lima 1974), cabbage (Hall & Smith 1983) and broccoli (Hamamoto & Mazelis 1986) as well as in turnip (*B. rapa*) (Anderson & Thompson 1979). The *B. oleracea* enzyme degrades L-cystine by an β-elimination reaction to form cysteine persulphide, pyruvate and ammonia (Hamamoto & Mazelis 1986).

1.2.3.2.2.d. Purification

Due to its pivotal role in flavour release, alliinase has been of great interest to researchers studying the flavour biochemistry of *Allium* species. Early preparations of alliinase from *A. sativum* and *A. cepa* by Stoll & Seebeck (1951) and other workers were relatively crude, but the high abundance of alliinase enzymes in *A. sativum* cloves and *A. cepa* bulbs (12% and 6% by weight of the soluble proteins respectively) (Nock & Mazelis 1987) enabled many of their characteristics to be determined. Attempts to isolate apparently homogeneous alliinase were eventually successful in the late 1970s when purifications and properties were independently reported for both the *A. cepa* and *A. sativum* enzymes (Tobkin & Mazelis 1979, Kazaryan & Goryachenkova 1978).

A. cepa and *A. sativum* alliinases are unstable, and generally lose activity at 4°C within hours of being isolated. However, in the presence of 10% glycerol (*A. sativum*) or 30% ethane diol (*A. cepa*), they can be stored for several weeks with no decline in activity (M.L. Shaw pers. comm., Jansen *et al.* 1989a) and these polyalcohol solvents are a feature of alliinase purification methods. Reducing agents such as β-mercaptoethanol, are also commonly used. The purification of *A. sativum* alliinase requires the presence of pyridoxal phosphate in the extraction buffers to maintain activity (Kazaryan & Goryachenkova 1978) and pyridoxal phosphate also enhances the stability of the *A. cepa* enzyme (M.L. Shaw pers. comm.). Otherwise the purification of both *A. cepa* and *A. sativum* alliinases is by standard precipitation and chromatographic separation methods.

1.2.3.2.2.e. Physicochemical characteristics

Determination of the holoenzyme size of purified alliinase from both *A. cepa* and *A. sativum* has been complicated by the tendency of these enzymes to aggregate in solution. Using a range of approaches, the *A. cepa* holoenzyme has been estimated as having a molecular mass of between 127 and 200 kDa with two, three or four equal subunits of usually approximately 50 kDa in size (see Table 1.1). The enzyme from *A. fistulosum* has also been estimated to be a 220 kDa tetramer composed of identical 52 kDa subunits (Fujita *et al.* 1991). The *A. sativum* enzyme has been described as having a molecular mass of 130 kDa with two subunits of 65 kDa (Kazaryan & Goryachenkova 1978) or a molecular mass of approximately 85 kDa with two subunits of 42 kDa (Nock & Mazelis 1986). Jansen *et al.* (1989a) more recently estimated that the *A. sativum* alliinase holoenzyme is a dimer of between 108-110 kDa with a subunit mass of 53.5 kDa. Consequently the *A. sativum* enzyme is most likely to be a dimer, also with subunits of approximately 50 kDa. An explanation which clarifies the confusion over the holoenzyme state has been proposed by Lancaster & Boland (1990). It states that the *A. cepa* and *A. sativum* alliinase subunits may aggregate into dimers, trimers or tetramers and therefore the native form observed is a time-averaged rapid equilibrium between these options. Nock & Mazelis (1987) found that the *A. cepa* enzyme is more difficult to dissociate into subunits than the *A. sativum* enzyme. This was thought to indicate differences in their subunit interactions.

Table 1.1. Alliinase holoenzyme and subunit size estimates

Method of Determination	Size (kDa)			
	A. cepa alliinase		*A. sativum* alliinase	
	Native	Subunit	Native	Subunit
Analytical ultracentrifugation	127[a]	85[a]		47[a]
Density gradient ultracentrifugation	150[b]			
Gel chromatography	140[a], 200[d]	50[d]	144[a], 130[c], 85[e], 110[f]	65[c], 42[e], 53.3[f]
SDS gel electrophoresis		53[a], 54[a], 50[b]		51[a], 49[a]

[a] Lancaster & Boland 1990. [b] Tobkin & Mazelis 1979. [c] Kazaryan & Goryachenkova 1979.
[d] Nock & Mazelis 1987. [e] Nock & Mazelis 1986. [f] Jansen *et al.* 1989a.

The related *Brassica oleracea* enzyme has been reported to be a trimer with subunits of 49 kDa (Hamamoto & Mazelis 1986). Mazelis & Creveling (1975) have estimated that the size of the *Acacia farnesiana S*-alk(en)ylcysteine lyase holoenzyme is 144 kDa, whilst the subunit has been found by denaturing gel electrophoresis to be 42 kDa (Sweet & Mazelis 1987). The related *Pseudomonas putida* alk(en)ylcysteine lyase holoenzyme was found to have a molecular mass of about 195 kDa and is thought to be composed of six identical subunits of approximately 37 kDa (Kamitani *et al.* 1990).

Most alliinases that have been purified have been shown by periodic acid-Schiff base staining to be glycoproteins. Nock & Mazelis have estimated that carbohydrate comprises between 4.6% (Nock & Mazelis 1987) and 5.8% of the weight of the *A. cepa* enzyme (Nock & Mazelis 1986). The *A. sativum* alliinase has been estimated to have a similar carbohydrate content of 5.5% (w/w) (Nock & Mazelis 1986) and the enzyme from *Brassica oleracea* was shown to contain 5.8% (w/w) carbohydrate (Hamamoto & Mazelis 1986). The oligosaccharides of *A. cepa* alliinase were found to be composed predominantly of mannose (78%), and also of galactose and glucose. In comparison, the side-chains of the *A. sativum* enzyme were found to contain 43% mannose and 34% glucose. The higher mannose content of the *A. cepa* alliinase is consistent with its greater affinity to concanavalin A Sepharose (Nock & Mazelis 1987). There were small amounts of arabinose (9%) and xylose (5%) detected within the glycans of the *A. sativum* enzyme, but these sugars were detected only at the 1% level within the *A. cepa* enzyme side chains (Nock & Mazelis 1987).

Both *A. cepa* and *A. sativum* alliinase have been reported to show some charge variation on isoelectric focusing gels, yielding multiple bands (Nock & Mazelis 1987). This was interpreted as evidence of carbohydrate micro-heterogeneity, and the major bands were visualized at pH 8.0 for the *A. cepa* alliinase and pH 7.0 for that of *A. sativum* (Nock & Mazelis 1987). However in a subsequent study the isoelectric point of *A. sativum* alliinase was found to be 4.9 and no charge heterogeneity was observed (Jansen *et al.* 1989a). It appears that more studies with highly purified protein will be needed to clarify this point.

The immunological cross-reactivity of *A. cepa*, *A. sativum* and *A. porrum* alliinases has been investigated using polyclonal antibodies raised to apparently homogeneous enzyme prepared from each species. Polyclonal antibodies raised to *A. cepa* alliinase reacted inconsistently to the *A. sativum* enzyme as determined by Ouchterlony double immunodiffusion, whereas polyclonals raised to the *A. sativum* enzyme showed no reaction to the *A. cepa* alliinase. There was no reactivity of either *A. cepa* or *A. sativum* anti-alliinase polyclonal antiserum with *A. porrum*

alliinase (Nock & Mazelis 1989), suggesting that these enzymes are immunologically distinct. However this low cross-reactivity may reflect differences in the carbohydrate moieties of these enzymes rather than in the epitopes presented by their protein back bones. These same authors determined the N-terminal amino acid sequence of the *A. sativum* enzyme to be lysine - methionine - threonine - tryptophan, but were unsuccessful in sequencing the N-terminus of the *A. cepa* enzyme.

1.2.3.2.2.f. Catalytic characteristics

In 1962 Schwimmer & Guadagni demonstrated that a highly significant correlation exists between the release of flavour compounds from *A. cepa* and the accumulation of pyruvate. Quantitative analysis of alliinase activity can be carried out by determining the substrate-dependent rate of pyruvate production, and this can be followed spectrophotometrically by measuring the amount of a 2,4-dinitrophenylhydrazine derivative of pyruvate produced over a course of time (Schwimmer & Weston 1961). An alternative assay couples alliinase activity to that of lactate dehydrogenase (LDH) so that pyruvate formed by the alliinase-catalysed reaction is reduced to lactate by excess LDH with the concomitant oxidation of NADH. The second reaction is stoichiometric with pyruvate formation and is followed spectrophotometrically by monitoring NADH oxidation at a wavelength of 340 nm (Schwimmer & Mazelis 1963).

Many α-amino acid-metabolizing enzymes are known to require the cofactor pyridoxal phosphate (PLP). PLP is essential for the activity of all alliinase and alliinase-like enzymes purified to date. A wide range of PLP antagonists such as aminooxyacetate, sodium cyanide, aminooxypropionate, D-cycloserine, L-cycloserine, 3-aminooxy-DL-alanine and 3-cyano-L-alanine inhibit alliinase activity from *A. cepa* and *A. sativum* (Tobkin & Mazelis 1979, Jansen *et al.* 1989b, Lancaster & Boland 1990). Reversible inhibition by hydroxylamine, a compound specific for PLP-requiring enzymes, has also been demonstrated for both enzymes. Furthermore, the extinction spectra of both *A. cepa* and *A. sativum* enzymes reveal a characteristic PLP absorption peak at 430 nm (Mazelis & Crews 1968, Nock & Mazelis 1986, Nock & Mazelis 1987). It has been estimated that the *A. cepa* enzyme contains three PLP moieties per 150,000 molecular weight, suggesting one is bound per equivalent alliinase subunit (Tobkin & Mazelis 1979). Early determinations of the PLP requirement of the *A. sativum* enzyme suggested three moles were bound to each equivalent subunit. However, this high estimate was

attributed to non-specific binding, as to maintain enzyme activity PLP was present in all the extraction buffers (Kazaryan & Goryachenkova 1978). Nock & Mazelis (1986) subsequently demonstrated that the *A. sativum* enzyme bound one mole of PLP per subunit molecule. Mazelis & Creveling (1975) found that the alliinase of *Acacia farnesiana* also displays this same cofactor:subunit ratio.

The β-elimination of the *S*-alk(en)yl sulphoxide group catalysed by alliinase is believed to occur via a PLP-Schiff base derivative (Block 1985). Sodium borohydride reduces a PLP-Schiff base to a stable secondary amine bond but does not interact with free PLP (Fischer *et al.* 1958). This reagent can be radiolabelled and used to select peptides containing PLP attachment sites from protein fragments generated by tryptic digestion. The binding site is then identified by peptide sequencing (Fischer *et al.* 1958). This procedure has been applied to a variety of pyridoxal enzymes, but attempts to trap PLP-Schiff base intermediates of alliinase using ^{14}C-labelled *S*-methyl-CSO and cyanoborohydride have to date been unsuccessful. The formation of a *S*-propyl cysteine-PLP derivative can be achieved using the alliinase substrate analogue *S*-propyl cysteine, and labelling studies with tritiated borohydride followed by hydrolysis have shown that an internal Schiff base is formed with the 4-amino group of a lysine residue (Lancaster & Boland 1990). Therefore it is probable that β-elimination via a PLP-Schiff base intermediate is the reaction mechanism of *A. cepa* alliinase. The determination of the active site peptides of alliinase will require further efforts to trap a PLP-Schiff base intermediate.

Addition of β-mercaptoethanol to *A. cepa* alliinase preparations is beneficial in retaining activity whereas the addition of 25 mM *N*-ethylmaleimide (a sulphydryl reagent) has no effect (Tobkin & Mazelis 1979). The prevention of a disulphide linkage was therefore suggested to be important in maintaining enzyme activity (Tobkin & Mazelis 1979), however the functional significance of this is not yet understood. Jansen *et al.* (1989b) observed that the activity of alliinase extracted from *A. sativum* cloves was inhibited by rotenone, a compound which inhibits the respiratory electron transport chain. Rotenone is thought to interact with the oxidized coenzyme Q and to inhibit the reduction of quinone to hydroquinone by a flavoprotein, and/or to interact with the electron transport-dependent 'phosphorylation site 1' by interfering with redox reactions during formation of ATP. These authors also observed that NADH produces a trough in the absorption spectrum of alliinase at 430 nm, which is characteristic of NADH-reduction of flavoproteins. The size of this trough was increased by the addition of rotenone, and together these results suggest that in addition to PLP, alliinase may require a flavine

prosthetic group (Jansen *et al.* 1989b). Although Block (1985) has previously proposed a metal ion requirement in the active site, Jansen *et al.* (1989b) could find no dependence or modification of *A. sativum* alliinase activity by the addition to 1 mM of Ba^{2+}, Ca^{2+}, Co^{2+}, Fe^{2+}, Mg^{2+}, Mn^{2+}, or Zn^{2+}, or by the addition to 1.5 mM of the chelating agent EDTA.

As is apparent from the range of precursor molecules discussed in Subsection 1.2.3.1.2, alliinase can cleave a variety of *S*-alk(en)yl-CSOs. Numerous studies have investigated the kinetics of both the *A. cepa* and *A. sativum* enzymes and activities have been found to vary with each of the naturally occurring and synthetic substrates. Although both (+) and (-) configurations of the sulphur oxygen are substrates, the naturally occurring (+) isomers are split much more rapidly (Whitaker 1976). The K_m (Michaelis constant) value for *A. sativum* alliinase with the natural diastereomer of *S*-allyl-CSO was found to be 0.5 mM (Kazaryan & Goryachenkova 1978). However most investigations have used synthetic *S*-alk(en)yl-CSOs that have a racemic sulphur atom (50% (+) and 50% (-)), hence the cleavage rates observed are a combination of the rates at which the two forms are utilized. The Michaelis constants (K_m values) that have been determined using various substrates are summarised in Table 1.2. From these data it appears that alliinase shows a lower affinity to *S*-methyl-CSO than to the other possible substrates tested.

Table 1.2. Michaelis constants determined for the alliinase enzymes of *A. cepa* and *A. sativum* with synthetic (±)-*S*-alk(en)yl-cysteine sulphoxide substrates

Substrate	*A. cepa* enzyme K_m (mM)			*A. sativum* enzyme K_m (mM)		
S-Allyl-CSO				2.2[g]	3.3[e]	6.0[c]
S-1-Propenyl-CSO	6.0[d]					
S-Propyl-CSO	3.8[b]	4.0[a]	11.0[d]	1.1[g]	3.0[c]	
S-Methyl-CSO	16.6[b]	34.0[d]		15.0[c]		
S-Ethyl-CSO	5.7[b]	9.4[g]		4.9[g]	5.7[e,f]	6.0[c]
S-Butyl-CSO	4.7[b]			5.0[c]		

[a] Schwimmer & Mazelis 1963.
[b] Schwimmer *et al.* 1964.
[c] Mazelis & Crews 1968.
[d] Schwimmer 1968 - this author states that the synthetic substrates used were all (+) isomers.
[e] Nock & Mazelis 1986.
[f] Nock & Mazelis 1987.
[g] Jansen *et al.* 1989a.

Inhibitor studies have shown that both *S*-alkylated and non-*S*-alkylated derivatives of unoxidized L-cysteine, as well as homologous α-amino acids such as DL-homocysteine, effectively act as substrate analogues of both the *A. cepa* and *A. sativum* alliinases (Schwimmer *et al.* 1964, Mazelis & Crews 1968, Jansen *et al.* 1989b). The functional reaction of both enzymes seems to require that the substrate is a derivative of (-)L-cysteine, the sulphur atom is linked to an aliphatic group, an unmodified amino acid group is present (probably for the formation of a Schiff base with PLP), and that the sulphur atom is in the sulphoxide form. It is believed that differences in active site configurations of the two enzymes exist, as the *A. cepa* enzyme can lyse molecules with an aryl substituent on the sulphur atom whereas the *A. sativum* enzyme cannot (Whitaker 1976).

Nock & Mazelis (1987) directly compared the enzymatic and physical properties of the *A. sativum* and *A. cepa* alliinase enzymes, and found that although the V_{max} values were almost identical with the synthetic substrate *S*-ethyl-CSO, there was a slight variation in the K_m values (5.7 mM and 9.4 mM respectively), supporting previous evidence that the active site configurations may differ. The characteristic absorption peak of the *A. sativum* enzyme at a wavelength of 430 nm is not observed in the presence of L-cysteine, and this is believed to be due to the formation of a thiazolidine ring from L-cysteine and enzyme-bound PLP, with an absorbance maximum at 330 nm. However this change could not be induced in the absorption spectra of the *A cepa* enzyme (Nock & Mazelis 1987). The pH optima also differ, with the peak *A. sativum* enzyme activity being detected at a pH of 6.5 (Mazelis & Crews 1968, Jansen *et al.* 1989a), while the pH optimum for the *A. cepa* alliinase was found to be between 7.6 and 8.0 (Schwimmer & Mazelis 1963, Tobkin & Mazelis 1979). The *Brassica oleracea* cystine lyase displayed a similar broad optimum between pH 8.0 and 8.5 (Hamamoto & Mazelis 1986). *A. sativum* alliinase was observed to optimally cleave *S*-allyl-CSO at a temperature of 33°C, and *S*-ethyl-CSO between 30°C and 35°C (Jansen *et al.* 1989a) consistent with a temperature optimum of 37°C previously determined with the substrate *S*-allyl-CSO (Stoll & Seebeck 1951). However, differences in the pH optima and temperature sensitivity of *A. sativum* alliinase are found when activity is assayed with *S*-propenyl-CSOs and *S*-methyl-CSO substrates, and this has led to a proposal that propenyl- and methyl-specific alliinase activities may exist in *A. sativum* (Lawson & Hughes 1991). This hypothesis is consistent with the tendency for alliinase to show a lower affinity to *S*-methyl-CSO than to the other substrates (see Table 1.2). Enzyme activity is completely destroyed by freezing *A. cepa* bulbs (Schwimmer & Guadagni 1968), but the mechanism of this inactivation is not known.

1.2.3.2.2.g. Alliinase distribution, location, and putative function

The rapidity of flavour release upon cell rupture led Becker & Schuphan (1975) to postulate that alliinase and its substrates were physically separated. Cell fractionation studies of protoplasts from *A. cepa* bulb inner scales later confirmed this, and demonstrated that the substrate *S*-alk(en)yl-CSOs are contained in the cytoplasm while alliinase is compartmentalized in the vacuole (Lancaster & Collin 1981).

Few reports on the expression of alliinase in different tissues or stages of development are available. That most tissues of *A. cepa* and *A. sativum* release flavour compounds when crushed is indirect evidence of the presence of alliinase (Lancaster & Boland 1990). There is little flavour release from the cortical pith tissues of the leaf blades and scape, however whether this is due to low levels of alliinase or of the flavour precursors is unclear (Lancaster & Boland 1990). Becker & Schuphan (1975) found by histochemical means that alliinase was present in a wide range of *A. cepa* bulb cells. Freeman (1979) observed that the activity of alliinase in *A. cepa* seeds was less than 2% of the level measured in bulbs, but that activity was readily detectable within a few days of germination, and subsequently increased to a maximum after 15 to 20 days. Redifferentiated *A. sativum* cultures were found to contain levels of alliinase activity much higher than in undifferentiated callus tissue (Malpathak & David 1986).

Alliinase is abundant in *A. cepa* and *A. sativum* tissues, representing up to 6% (w/w) of the soluble proteins of *A. cepa* bulbs and 12% (w/w) of the soluble proteins of *A. sativum* bulbs (Nock & Mazelis 1987). Why this enzyme is accumulated to such concentrations is unknown. Extracts of *A. cepa* and *A. sativum* are known to have an anti-microbial effect on many plant pathogens, and the antimicrobial effect of diallyl thiosulphinate (allicin) is considerable (reviewed by Virtanen 1965, Fenwick & Hanley 1986). The possible role of flavour release as a defence mechanism is consistent with the comparatively few insect pests of *Allium* species that are known. The occurrence of some insect and fungal pests which have developed highly host-specific, specialized interactions with *Allium* species also supports this hypothesis (discussed in Lancaster & Boland 1990). Mazelis & Fowden (1973) found that *Acacia georginae* and *Acacia julibrissin* seedlings showed much higher alliinase specific activity in the hypocotyl than in the cotyledon or shoot tissues, which led these authors to postulate that alliinase-mediated release of volatile sulphide compounds during early development may function in defence against soil-borne pathogens and facilitate seedling establishment.

1.3. Plant gene structure and expression

1.3.1. Overview

In recent years a considerable number of higher plant protein-encoding genes have been isolated. Recently, Wada *et al.* (1992) listed nine dicot and four monocot species for which more than 20 independent protein-encoding gene sequences have been entered in the GenBank database. The rapidly increasing number of known DNA sequences has allowed comparisons of gene structure from different plant classes to be made. These compilation analyses have demonstrated slight differences between the structure of monocot and dicot genes, for example monocot genes tend to be more G+C-rich than dicot genes in their coding regions (Murray *et al.* 1989, Campbell & Gowri 1990) and do not exhibit the strict requirement for A+T-rich introns shown by dicot transcripts (Goodall & Filipowicz 1991). Differences in translation initiation consensus site contexts between these two plant groups have also been found (Cavener & Ray 1991). However, it is evident from the data compiled by Murray *et al.* (1989) and Campbell & Gowri (1990) that analyses of monocot gene base composition tend to be highly biased by the predominance of genes from the four cereal species maize, barley, wheat, and rice, and there is little evidence that the propensities demonstrated by the gene sequences from these graminaceous plants are exhibited by the transcribed sequences of all monocots. A general conclusion that can be drawn from the accumulated data is that plant genes are very much like animal or yeast genes in structure, and the essential structural features of protein-encoding plant genes appear to be consistent throughout all plant classes. This section is devoted to a discussion of plant gene structure and the relationship between gene structure and gene expression.

Plants are morphologically simple organisms with less than 60 cell and tissue types. Despite this morphological simplicity a large number of mRNA species may be expressed in any organ. For example, Goldberg (1988) estimated that the cells of tobacco contain approximately 25 000 different mRNA types, about 6 000 to 10 000 of which are unique to specific plant organs. Additionally, each cell or tissue type in an organ has a distinct structural or physiological role, and as a consequence may express different sets of genes. In many instances the production of a polypeptide may only occur under certain conditions or in specific cell types within the plant. Cellular processes that specifically regulate thousands of genes in response to different developmental situations and environmental cues must be present. It has

been shown that the production of polypeptides can in some cases be controlled at the level of mRNA or protein stability, but for most genes expression is primarily controlled at the level of transcription. Mechanisms by which protein production is controlled by differential plant gene expression are now beginning to be understood.

In higher organisms, proteins are frequently encoded by gene families whose members are closely related. These genes can encode polypeptides that are structurally indistinguishable, but whose expression may be independently or differentially controlled. The well studied ribulose biphosphate carboxylase small subunit (*rbcS*) gene families encode functional translation products which show little or no amino acid sequence heterogeneity within all plant species where more than one gene sequence has been obtained. However, these related genes can show highly divergent expression characteristics (Dean *et al.* 1989b). In petunia, *rbcS* gene expression varies quantitatively, ie. there are highly expressed genes and lowly expressed genes, but all *rbcS* genes appear to be expressed in the same organs at the same relative levels and at the same time in development. In tomato there are differences in the transcript levels and the patterns of expression (Dean *et al.* 1989a). It is often of value to separately study the expression of one or several members of a gene family. The establishment of reproducible plant transformation procedures, via *Agrobacterium*-mediated transformation (Horsch *et al.* 1985, Klee *et al.* 1987, Hooykaas 1989), DNA uptake into protoplasts (Draper *et al.* 1982, Gasser & Fraley 1989, Davey *et al.* 1989) and high velocity microprojectile bombardment of plant tissue (Klein *et al.* 1987) have provided a means of manipulating gene expression and identifying the multiple *cis*-acting sequence elements thought to be involved in gene regulation.

1.3.2. Gene families

The fundamental characteristics of multigene families are multiplicity, close linkage, sequence similarity and overlapping function (Hood *et al.* 1975). Multigene families range in size from two or three genes per haploid genome to complex superfamilies, for example the petunia actin gene family which contains in excess of 200 copies (Baird & Meagher 1987). Large gene families can often be classified into subfamilies on the basis of variations in sequence homology. This has been done with the zein gene family, where the sequence divergence between the four subfamilies is known to range from 15 to 35% (Heidecker & Messing 1986). The existence of gene families was originally postulated from protein sequencing data, but their ubiquity has since been most convincingly demonstrated by recombinant DNA techniques. For example, at least five petunia chalcone synthase (CHS)

isopeptides with differing molecular weights and isoelectric points have been shown to be the products of a gene family with between six to eight members (Ryder *et al.* 1987). Isozymes with highly divergent functions are known, such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH) for which both cytoplasmic and chloroplastic forms exist (Martinez *et al.* 1989, Russel & Martin 1989). The cytoplasmic form is involved in glycolysis, whereas the chloroplast enzyme catalyses the reverse reaction in the Calvin cycle. The characterization of these isozymes has been greatly enhanced by the isolation of their clones, which has shown that the chloroplastic enzyme in maize is a heterodimer encoded by the *gapA* and *gapB* genes, while the cytoplasmic enzyme is homomeric and encoded by several *gapC* genes.

Molecular analysis of gene family members also allows the identification of DNA sequence variations that do not result in changes at the protein level, such as within the previously mentioned *rbcS* gene families of several species which encode structurally indistinguishable polypeptides (Dean *et al.* 1989b). Sequence divergence among members of a gene family tends to be greatest within non-coding regions, presumably because selection operates on the basis of protein structure (Li *et al.* 1985). Dean *et al.* (1985) has demonstrated that the greatest degree of sequence variation of transcribed gene family sequences often occurs just 3' to the translation termination codon. Nucleotide changes that do not alter the protein-coding properties of the mRNA are also frequently found at the third base of many codons (Hood *et al.* 1975).

1.3.2.1. Function of gene families

Proteins that are required rapidly and in high abundance, such as the storage proteins β-conglycinin (Harada *et al.* 1989), glycinin (Nielsen *et al.* 1989), legumin (Heim *et al.* 1989), patatin (Mignery *et al.* 1988), zein (Heidecker & Messing 1986), and many others, have been found to be encoded by multigene families. Expansion from a single copy gene into a multigene family is believed to answer the plant's need to rapidly produce large amounts of a specific protein (Heidecker & Messing 1986). The development of a gene family also effectively releases gene copies from the selective pressures acting on the original gene. Multiple gene copies encoding a protein with an important cellular function may allow one copy of the gene to mutate, enabling a variant protein with a potentially advantageous function or expression to be generated without the risk of the mutation being lethal to the organism (Li *et al.* 1985).

The existence of pseudogenes, i.e. sequences with no physiological value,

supports the concept that mutations occur. Pseudogenes show high sequence similarity to corresponding active genes, but contain defects that prevent the formation of a functional translation product. This can result from small alterations to the gene such as a single base deletion, or can be due to major deletions which halt RNA transcription. For example, the gene coding for the phytohemagglutinin E polypeptide from the Pinto cultivar of *Phaseolus vulgaris*, *Pdlec1*, was found to contain a single base pair deletion in codon 11 which caused premature termination of translation. Interestingly, this gene was thought to be transcribed normally, but the transcripts showed a short half-life in the cytoplasm which could not be explained (Voelker *et al.* 1990). The existence of pseudogenes appears to be a widespread phenomenon, and they have been found in most multigene families which have been subjected to extensive DNA sequencing (Li *et al.* 1985). Pseudogenes can represent a significant proportion of a gene family, for instance up to 50% of the maize zein multigene family is comprised of functionally inactive sequences (Heidecker & Messing 1986).

1.3.2.2. Expression of gene family members

Individual members of multigene families often exhibit highly divergent expression patterns. Inducible expression of one or more members of a multigene family in response to developmental or environmental cues is very widespread and has been well documented among the members of many plant gene families e.g. the actin gene families of petunia (Baird & Meagher 1987) and rice (McElroy *et al.* 1990a) and the petunia *rbcS* gene family (Dean *et al.* 1985). Hence regulatory elements present in the genes or their flanking regions must differ for different members of gene families. A dramatic example of differential gene family member expression is found among the maize genes encoding GAPDH. The *gapA* and *gapB* genes which code for the chloroplastic forms of GAPDH are induced by light, whereas *gapC* genes for cytosolic GAPDH are not light-induced. Genes for glycolytic enzymes are usually induced during anaerobiosis, because the low energy efficiency of fermentation requires an increased flux through the glycolytic pathway. Of the cytosolic GAPDH genes, only *gapC1* is anaerobically induced, whereas *gapC2* mRNA levels remain constant (Martinez *et al.* 1989, Russel & Martin 1989).

The range of expression patterns exhibited by different members of plant gene families suggests this flexibility confers some evolutionary advantage. Gene families may provide a means of limiting the synthesis of a certain protein to specific tissues or developmental stages, or in response to certain environmental conditions, and thus may enable the plant to react in a highly adaptable and efficient way.

1.3.3. Genomic plant gene structure

1.3.3.1. Transcriptional control signals

The synthesis of a functional mature polypeptide results from the interaction of multiple control factors. Processes thought to be potentially involved in the regulation of gene expression include nuclear RNA modification, splicing and turnover, selective RNA transport from the nucleus to the cytoplasm, mRNA turnover in the cytoplasm, translation, post-translational processing, compartmentalization and protein turnover (Goldberg *et al.* 1989). However, control of gene expression in most cases has been found to be primarily at the level of transcription.

Transcription of protein-encoding genes in eukaryotes requires a complex series of interactions involving RNA polymerase II, additional transcription factors and several specific nucleotide sequences positioned at the 5' end which function in the initiation and regulation of transcription. These *cis*-acting elements have been studied in some depth and are common to many plant genes. The identified sequence motifs fall into two categories. The TATA box, or a functionally related sequence, binds the RNA polymerase complex and determines the site at which transcription will initiate. A TATA box is probably present in all functional genes. The second type of sequence motifs are those found in certain genes whose presence can sometimes be correlated with particular expression patterns, or can be experimentally demonstrated to mediate a specific expression pattern in response to a particular signal. In this review the major *cis*-acting recognition signals which have been found within nuclear plant gene sequences will be discussed. A review of the sequence elements that function in RNA processing and translation will follow this. Rather than provide an exhaustive overview, the focus is on relevant studies that are intended to illustrate the current understanding of the role of *cis*-acting sequence elements and the *trans*-acting binding factors.

1.3.3.1.1. The TATA and CAAT box signals

The most commonly found *cis*-acting sequence, the TATA box, is a short A+T-rich region located 5' to the transcription start site. The sequence of the TATA box is conserved among many eukaryotic genes. This region plays an important role in promoter activation and in specifying transcription start sites (Breathnach & Chambon 1981). Unless the TATA box has been replaced by sequences that can substitute functionally (Chen & Struhl 1989), deletion of this element leads either to zero or drastically reduced levels of transcription (Chen & Struhl 1988).

Binding of the transcription initiation factor (TFIID) to the TATA box is the first step in the initiation of transcription. TFIID is thought to recruit the general factors TFIIA, TFIIB, TFIIE and TFIIF, as well as RNA polymerase, to form a transcription complex through stable binding to the TATA box (Pugh & Tjian 1990). Additional transcription factors responsible for regulated transcription are thought to help assemble and/or engage the transcription complex. These may bind to other *cis*-acting sequences and interact with TFIID directly to stabilize its binding to the TATA box. Alternatively these factors may function via co-activator proteins that have no affinity for DNA themselves but are thought to have contact sites for both RNA polymerase and upstream DNA-binding proteins (Lewin 1990, Wefald *et al.* 1990). The plant gene consensus for the TATA box and surrounding region has been shown to be TCACTATATATAG, and was found unaltered in 84% of the 79 plant genes examined by Joshi (1987a).

Plant gene promoters generally direct transcription from a single initiation site located 32 ± 7 nucleotides downstream of the TATA box (Joshi 1987a), suggesting this specific spacing is strictly required. Examples have been found of plant genes that may use more than one transcription start site, resulting in transcripts with varying 5' untranslated region lengths. For instance, RNase protection analysis indicated up to four possible sites could function as the site of transcription initiation for a lectin gene isolated from *Ricinus communis*, and this may have resulted from the presence of at least two potential TATA boxes located in the 5' flanking region (Tregear & Roberts 1992). This redundancy of TATA boxes is observed within the 5' flanking sequences of several plant genes which initiate transcription from more than one point. It has been suggested that developmental or environmental cues may determine which transcription start site is functional (Bird *et al.* 1988).

There is some evidence that regions flanking the TATA box can be important in regulating gene expression. Expression analysis of promoter mutants of a maize gene which encodes alcohol dehydrogenase (*adh1*), indicates that organ-specific and possibly stage-specific expression of this gene can be controlled by sequences in the vicinity of the TATA box (Kloeckner-Gruissem *et al.* 1992). Further evidence has also been provided by the analysis of promoter regions of several genes encoding the seed storage protein napin from *Brassica napus*. It was found that these genes contain two consensus motifs, TACACAT and CATGCA, which frequently occur as overlapping, direct repeats. One copy of the CATGCA motif occurs in close proximity to the TATA box, while the TACACAT sequence was found to interact with a protein present in crude nuclear extracts from *B. napus* seeds (Ericson *et al.* 1991).

A second conserved region believed to be involved in transcription regulation is the sequence CCAAT, which occurs in many animal gene promoter sequences at a position between 70 and 80 nucleotides proximal to the transcription start site (Nevins 1983, Dynan & Tjian 1985). This sequence has been observed within the promoter region of some plant genes. For example, the nopaline synthase gene has a CAAT site 80 nucleotides upstream from the cap site (Messing *et al.* 1986), the tomato polygalacturonase gene contains two putative CAAT boxes 69 and 62 bases 5' of the major transcription start site (Bird *et al.* 1988), and the petunia *cab* genes encoding the chlorophyll *a/b* binding protein contain this sequence motif 50 to 55 nucleotides upstream of the TATA box (Dunsmuir 1985). However, the position of this sequence motif can vary, for instance Shinshi *et al.* (1990) found the sequence CCAATT at position -1114 relative to the transcription start site of a tobacco endochitinase gene. Many plant genes appear to lack a consensus CAAT box (eg. Hattori & Nakamura 1988, Lois *et al.* 1989), and whether it plays a significant role in plant gene expression is as yet unclear.

1.3.3.1.2. Enhancers and silencers

Enhancers are defined as sequence elements that can act as long-range activators of gene transcription (Serfling *et al.* 1985). These motifs lack promoter activity and are unable to direct transcription themselves but can increase transcription from promoters. In mammalian systems, enhancers are often found relatively distant from the site of transcription initiation and are known to function when moved to other sites. Enhancer-like elements have been found in the flanking sequences of several plant genes, predominantly in the 5' regions. For example, upstream elements of the cauliflower mosaic virus (CaMV) 35S promoter have been shown to induce expression from weaker promoters such as a truncated *rbcS-3A* promoter (Fang *et al.* 1989). An enhancer-like sequence element, GTGGTTC was located between 263 and 269 bp 5' to the transcription start site of the parsley CHS gene (Herrmann *et al.* 1988), and Kaulen *et al.* (1986) found the similar sequence GTGGTTAG, which is identical to the consensus core sequence for enhancers in animal genes, within the region 564 bp to 661 bp upstream of the *Antirrhinum majus* CHS gene transcription initiation site. The petunia *rbcS* gene, *SSU301*, has sequences at both the 5' and 3' regions which affect its transcription, and a 81 bp fragment from the 5' flanking region of this gene was described as having an 'enhancer-like' activity (Dean *et al.* 1989a).

The presence of polypurine/polypyrimidine sequences within the promoters of yeast and mammalian genes has been noted by many groups, and these sequences

have also been found in the 5' regions of several plant actin genes (Pearson & Meagher 1990). These (dA-dT) tracts are thought to act as enhancers, and longer tracts have been found to stimulate transcription more effectively than short ones (Chen *et al.* 1987). Wang *et al.* (1992) have shown a 38 bp poly(dA-dT) sequence to be a positive regulator of the expression of a rice actin gene (*act1*) in transformed maize and rice protoplasts. A poly(dA-dT) binding protein was also shown to be associated with this activity. Transient expression assays in transformed rice protoplasts of *act1* promoter deletions also identified a CCCAA pentamer which acted as a negative regulator of expression (Wang *et al.* 1992).

The study of Rolfe & Tobin (1991) has provided evidence for silencer elements within the promoter regions of plant genes. Transient expression of deletion mutants of the *Lemna gibba rbcS* gene *SSU5B* showed that deletion of the 5' region to 301 bp upstream or to 480 bp upstream of the transcription initiation site had no significant effect, whereas deletion to 397 bp significantly inhibited transcription. This was interpreted as evidence of a silencer element between positions -301 and -397[1]. Simpson *et al.* (1986a) have similarly identified a silencer sequence within a 247 bp upstream region of a pea light-harvesting chlorophyll *a/b* binding protein (*lhcp*) gene which was able to repress the normally constitutive activity of the nopaline synthase promoter in roots of transgenic plants, and is therefore thought to contain a tissue-specific negative control element.

1.3.3.1.3. Sequence elements conferring specific gene expression patterns

A gene is considered to be constitutively expressed when the expression level remains constant on a per cell basis under all conditions examined. This assumes no variation of mRNA abundance and/or translational efficiency during development, in different organs, or under differing environmental conditions. As the analysis of gene expression becomes more sensitive, genes formerly believed to be constitutive are now being found to be regulated, and it is becoming difficult to identify a cellular plant gene that is unequivocally constitutive. Dehesh *et al.* (1991) have suggested that the *phyB* genes of rice and *Arabidopsis* may be constitutively expressed, as their transcript levels (unlike other phytochrome genes) have been shown to be unaffected by a range of light treatments. However, other possible control mechanisms have not yet been completely eliminated, and further analysis may yet reveal some regulatory mechanism. Most plant genes characterized to date are transcribed in a regulated rather than a constitutive fashion (Kuhlemeier *et al.*

[1]where +1 is the transcription start site and the nucleotides 5' of this are assigned negative numbers

1987). Expression of plant genes is frequently tissue-specific and/or stage-specific, and may be influenced by a range of environmental factors such as light, stress, hormones, or attack by pathogens.

It can be postulated that transcriptional gene regulation derives from interactions of different *trans*-acting factors with specific recognition sites in various *cis*-acting domains. The characterized *trans*-acting factors involved in the regulation of the CaMV 35S promoter and many plant genes show strong homologies, suggesting regulation of expression may result from the action of different combinations of common *trans*-acting factors (Sanfaçon 1992). Footprint analyses, deletion derivatives and introduction of promoter-reporter gene fusions into transgenic plant tissues and protoplasts, have greatly delineated the role of 5' *cis*-acting elements and *trans*-acting binding proteins in the responses of plants to certain stimuli (see Kuhlemeier *et al.* 1987, Edwards & Coruzzi 1990, Kuhlemeier 1992 for reviews). Palindromic sequences are a common, but not essential, feature of these *cis*-acting elements. This is thought to reflect the structure of proteins involved in sequence-specific protein-DNA interactions, which are often active as dimers or tetramers, and bind with a two-fold symmetry that matches the symmetry of the palindrome (Takeda *et al.* 1983).

Further to this, specific expression patterns may result from similar *cis*-acting elements appearing in different combinations in each gene promoter (Yamamoto 1985), thereby leading to dissimilar interactions with a given set of *trans*-acting factors common to all cell types (Bustos *et al.* 1991). There is great potential for combinatorial diversity in such systems, which may contribute to the diversity of responses. The finding that combinations of *cis* elements can produce expression patterns not generated when they are in isolation suggests that a combinatorial code directs expression throughout development.

The CaMV 35S promoter has been used as a paradigm for dissection of regulatory sequence elements and has clearly demonstrated the modular organization of transcriptional control in plants. Although the 35S promoter is originally from a plant pathogen, its expression depends entirely on the plant transcriptional machinery. In early experiments approximately 1000 bp of promoter DNA, including a few base pairs beyond the transcription start site, were fused to various reporter genes. Using protoplasts, transformed calli and transgenic plants, reporter genes were always observed to be expressed at high levels, and were insensitive to endogenous and environmental cues such as hormones, heat shock or light (Kay *et al.* 1987, Odell *et al.* 1985, Ow *et al.* 1987). This 'constitutive' expression led to the popular use of the 35S promoter in experiments analysing other, regulated,

promoters, and in plant transformation studies. Later dissection of the 35S promoter into subdomains showed these are able to confer tissue-specific patterns of gene expression. For example, when the region lying upstream of position -90 of the transcription start site was deleted, considerable levels of the fused chloramphenicol acetyltransferase (CAT) marker gene mRNA and enzyme activity could be detected in roots, but not in leaves or stems (Poulsen & Chua 1988). When selected subdomains are combined, they confer expression patterns not detected from the isolated subdomains, suggesting that the 35S promoter has a modular organization and that synergistic interactions occur among the different *cis* sequence elements (Benfey & Chua 1990).

On the basis of studies with the CaMV promoter Benfey & Chua (1990) have proposed two models for synergistic combinatorial activity of *cis* elements. The models assume the sample case of two *cis* elements that each interact individually with cell- or tissue-specific *trans*-acting factors. On their own these elements confer expression in a particular cell type yet they confer expression in a third cell type when combined. The first model proposes that a promoter containing a combination of binding sites can confer a particular expression pattern because heterologous factors will bind cooperatively to different *cis* elements when these factors are present in low concentration. Therefore a productive complex can be attained when any single binding factor is present at levels too low to activate transcription from promoters containing only one type of binding site. The second model postulates the involvement of a target factor that mediates the interaction between more than one heterologous binding factor and their cognate *cis* elements.

The modular organization of promoter regions has been supported by studies of many plant genes. For example, deletion analysis of the promoter region of the gene encoding phenylalanine ammonia lyase 2 (PAL2), has shown that functionally redundant elements located between 74 bp to 289 bp upstream of the transcription start site were essential for expression in the xylem, but did not have a role in the expression of PAL in leaf primordia, stem nodes, or petals. However, deletion from positions -289 to -135, or from -135 to -74 did not abolish expression in xylem tissue (Leyva *et al.* 1992). This study also found that the -135 to -119 region implicated in xylem expression contained a negative element that suppresses the activity of a cryptic *cis* element for phloem expression located between -480 and -289.

Transcription rate and its regulation are likely to be determined not just by the intrinsic properties of a transcription factor and its cognate binding site, but rather by a complicated interplay of multiple factors and multiple binding sites. One factor may have different affinities to multiple sites and it may bind cooperatively. Factors may also compete for single or overlapping binding sites, resulting in changed interactions with the RNA polymerase complex.

1.3.3.1.3.a. Sequence elements conferring light-regulated gene expression

Light is one of the most important environmental cues in plant development, and is known to induce the activity of over 40 enzymes (Tobin & Silverthorne 1985). Consequently, light-regulated gene expression is probably the most well studied transcriptional controlling mechanism of plant genes, being responsible for the level of transcription of both nuclear and chloroplastic genes. Deletion analysis of plant gene promoters has confirmed the presence of multiple *cis*-acting regions involved in light-regulation (see Kuhlemeier *et al.* 1987, Okamuro & Goldberg 1989, Gilmartin *et al.* 1990, Thompson & White 1991, Rolfe & Tobin 1991 for reviews). Many of these have not yet been fully characterized.

Among the most thoroughly characterized of the light-responsive elements are the regulatory sequences of the pea *rbcS-3A* gene. The *rbcS-3A* promoter region has multiple light responsive elements, the GT-1 boxes, upstream (boxes II*, III*, II** and III**) and downstream (boxes II and III) of position -170, which bind the nuclear factor GT-1 (Green *et al.* 1987, Green *et al.* 1988). This redundancy of *cis*-acting elements may facilitate differential gene expression. The box II core sequence, GGTTAA, has been shown to be required for GT-1 binding *in vitro* (Green *et al.* 1988). Similar sequence motifs are also present within several other light-responsive genes (Dean *et al.* 1989a, Dehesh *et al.* 1991, Gilmartin *et al.* 1990). However, the other GT-1 boxes show considerable variation from this sequence, suggesting that GT-1 may bind to many related sequences. Mutational analyses have shown differences in the ability of combinations of box II and box III to activate transcription, which correlated with an altered affinity for GT-1 binding *in vitro*. This suggested that the strength of GT1 binding to an element may influence the level of transcription (Gilmartin *et al.* 1990). There is an absolute requirement within a -170 bp deleted *rbcS 3A* promoter for two GT-1 boxes for transcriptional activity, and sequences outside of the *in vitro* defined core binding site are also required for transcriptional activity *in vivo* (Gilmartin *et al.* 1990). GT-1 binding is necessary, but not sufficient, for transcriptional activation and a requirement for an interaction between GT-1 and other proteins is thought to be necessary for the formation of a stable transcription initiation complex (Gilmartin *et al.* 1990).

The G-box, which is found in several light-responsive promoters, is an upstream element with the consensus sequence TCTTCCACGTGGCAYY[2] (Giuliano *et al.*

[2]where Y = pyrimidine (C or T)

1988, Schulze-Lefert *et al.* 1989b). The role of the G-box in binding nuclear proteins in response to light has been demonstrated by *in vivo* footprint analysis and genomic sequencing of the parsley CHS promoter region. This has revealed four light-inducible *in vivo* footprints which are assembled within two independently functional light-responsive units - boxes I plus II (unit 1), and boxes III plus IV (unit 2), which control gene expression in a synergistic manner when present in combination (Schulze-Lefert *et al.* 1989a, 1989b). A core of essential nucleotides in boxes II and III of CACGTGGC has been defined. This core sequence is essential for the functioning of each unit, but the second box within each light responsive unit did not appear to require a strictly conserved sequence, although the relative positions of the two boxes was found to be important (Block *et al.* 1990). This study supports the hypothesis that the functional capacity and specificity of some *cis*-acting sequences may be conferred by combinatorial diversity involving flanking partner elements.

A 52 bp region corresponding to unit 1 is sufficient to confer light-mediated regulation upon a heterologous TATA box (Weisshaar *et al.* 1991). Box II-like sequences, or ACGT-containing promoter elements, are not exclusively involved in photoregulation, as similar sequences are present in plant gene promoters which are responsive to environmental and developmental signals other than light (Schulze-Lefert *et al.* 1989b). Common plant regulatory factors (CPRFs) that interact with members of the ACGT-containing promoter elements have been identified and their cDNAs cloned. These proteins contain basic domain/leucine zipper DNA binding motifs and are dimers that recognize binding sites with dyad symmetry. These factors bind to different ACGT-containing promoter elements with markedly different affinities, and distinct ensembles of nuclear factors bind to boxes II and III, despite their core sequence identity (Armstrong *et al.* 1992). Sequences flanking the G-box may determine the specificity of binding by *trans*-acting factors, and the affinity of binding of G-box proteins for the different classes of G-box elements may regulate specific expression patterns (Williams *et al.* 1992).

1.3.3.1.3.b. Sequence elements conferring tissue-specific gene expression

A large number of expressed mRNAs may be unique to a particular organ or tissue. Anthers have been estimated to contain approximately 11 000 diverse and unique mRNAs, roots about 7 000, and stems about 6 000 (Goldberg 1986, Goldberg 1988). When organ-specific gene regulation is analysed by *in situ* hybridization or reporter gene fusions, tissue-specific or cell type-specific patterns of gene expression

are observed (Okamuro & Goldberg 1989). The *rbcS* (Dean *et al.* 1985) and *lhcp* genes (Simpson *et al.* 1986b) exhibit tissue specific gene expression, and it is the same region that contains the light-responsive elements that also confers expression in chlorophyll-containing tissues. Using deletion mutants, Kuhlemeier *et al.* (1988) have demonstrated that sequences between -410 to -171 and the 166 bp region upstream of the transcription start site of the pea *rbcS-3A* gene were sufficient to encode light inducibility and organ specificity.

The sequences thought to be responsible for organ-specific expression of seed protein genes have been studied in some depth. Embryo-specific regulatory elements directing expression of the soybean α-subunit of β-conglycinin have been analysed using mutants containing deletions spanning the region -8.5 to +14 kb relative to the transcription start site of CG-1. Chen *et al.* (1986, 1988) showed that only the 159 bp contiguous to the transcription start site of the CG-1 β-conglycinin gene are required for expression during seed development. Expression of this seed storage protein in transgenic petunia indicated, however, that the region between -159 to -257 was important in directing high levels of expression. Five repeats of the sequence AVCCCA[3] and two imperfect 28 bp repeats, each containing two copies of this 6 bp G-C rich sequence, were found between positions -257 and -131, and it was concluded that G-C rich repeats were critical in determining embryo-specific expression (Chen *et al.* 1986). The importance of this region was further confirmed by demonstrating that a 170 bp fragment containing the G-C rich repeats could enhance CaMV 35S promoter/CAT reporter gene expression 25-fold in an orientation independent manner (Chen *et al.* 1988). Expression was position dependent, and limited expression was observed when these repeats were placed 3' to the reporter gene. Sequence comparisons of the 5' flanking regions of CG-1 and a further soybean β-conglycinin gene family member, CG-4, revealed short sequences of homology, including the CATGCAT and CAACACA consensus sequences found in the 5' regions of several seed protein genes (Harada *et al.* 1989). This conservation suggests that these sequence elements may be important in regulating seed protein gene transcription levels during germination.

Multiple *cis*-acting DNA domains appear to be involved in determining both the location and timing of the appearance of a polypeptide. The transcriptional regulation of a β-phaseolin gene of bean is primarily spatially determined by two upstream activating sequences (UAS). UAS1 at position -295 to -109 was sufficient for seed-specific expression, while a second domain UAS2 at position -468 to -391

[3]where V = A or G or C

extended gene activity to the hypocotyl (Bustos *et al.* 1991). The sequence GATAAGGCTA, which has homology to the binding site of ASF-2, a factor identified as binding to the CaMV 35S promoter (Lam & Chua 1989), was identified as important in the shoot-specific and light-inducible expression of the rice gene GOS5 (de Pater *et al.* 1990). This sequence is able to confer leaf-specific expression when fused to the -90 derivative of the CaMV 35S promoter and it is also present within the promoters of several light-responsive genes (Lam & Chua 1989).

The tissue-specific and developmentally regulated expression of the 5-enolpyruvylshikimate-3-phosphate synthase gene (EPSPS) has also been well studied. In petunia EPSPS is highly expressed in the root and stem of seedlings, and in mature plants expression is high in the petal, lower in the floral tube, and barely detectable in root, stem, and leaf (Gasser *et al.* 1988). Four *cis*-acting regions are believed to be involved in the regulation of this gene. The EP1 element is located between -1170 and -1150 relative to the start of transcription, and contains the palindrome TGACAGTGTCA, while EP2 (at -1307 to -1292), EP3 (at -1366 to -1345) and EP4 (at -1425 to -1408) all contain the conserved sequence TTACT$(N)_3$AT[4] in different orientations. These sites are all associated with the DNA-binding factor EPF1. This *trans*-acting protein has a similar expression pattern to EPSPS and its upstream region confers a nearly identical cell-specific expression pattern when fused to a reporter gene (Takatsuji *et al.* 1991). Other *trans*-acting factors are also involved in binding to the *cis* elements and it is thought that simultaneous binding of several factors may be required to activate transcription (Takatsuji *et al.* 1991).

Negative control of tissue-specific regulation by specific sequence elements has also been reported. A detailed analysis of the cell-type specificity conferred to a GUS reporter gene by various petunia *chsA* promoter fragments in transgenic petunia plants showed that TACYAT repeats within the -67 to -53 region could act as dominant negative *cis*-acting modules in controlling organ-specific expression. In the presence of the TACYAT module the CaMV 35S enhancer region only enhanced GUS activity in those organs in which the *chsA* promoter was normally active (van der Meer *et al.* 1992a).

1.3.3.1.3.c. Sequence elements conferring expression in response to environmental cues

Environmental influences such as heat shock, water stress, infection by pathogens, plant growth regulators, and wounding, have been demonstrated to induce the transcription of specific RNAs. Run-off transcription experiments with nuclei

[4]where N can be either A, C, G or T

isolated from control and heat-stressed soybean plants have demonstrated that induction of the low molecular weight heat shock proteins occurs at least partly at the level of transcriptional initiation (Kuhlemeier *et al.* 1987). Two maize polyubiquitin genes whose expression is induced to higher levels at elevated temperatures, show the presence of two overlapping sequences, CTGGACCCCTCTCG and CTCGAGAGTTCCGC, at positions -214 and -204 relative to the transcription start site. These sequences are similar to the consensus heat shock element found in other heat inducible genes (Christensen *et al.* 1992). The 5' flanking region of a tobacco endochitinase gene also revealed a 6 bp imperfect inverted repeat sequence at position -140, which resembled the consensus heat shock regulatory elements of animals (Shinshi *et al.* 1990).

Disease resistance in higher plants is manifested by the accumulation of a number of host-synthesized polypeptides produced in response to pathogen attack. Mutant promoter analysis, expression of a heterologous promoter fusion, and trans-competition experiments have demonstrated that two different *cis* elements within the bean *chs15* promoter, the G-box (CACGTG) and the H-box ($CCTACC(N)_7CT$), in combination are necessary and sufficient for stimulation of this gene by the phenylpropanoid-pathway intermediate ρ-coumaric acid (4-CA). Levels of 4-CA correlate with the activity of CHS transcription after fungal elicitor treatment, suggesting 4-CA may act as a signal molecule during initial activation of *chs* gene transcription by pathogen attack (Loake *et al.* 1992).

1.3.3.2. Signal transduction

As considerable evidence on the internal and external signals modulating gene expression is now available, the intermediary steps have become a major topic of interest. In the case of light-mediated responses, some details of the initial changes that act to cause red/far-red reversible reactions are known. The physical conformation of the photoreceptor phytochrome can be altered by red light, and far-red light can reverse this change (Thompson & White 1991). As an example of this, wheat protoplasts respond to red light treatment by swelling, and this response is reversible by exposure to far-red light. This increase in volume requires Ca^{2+}, and can occur in the dark when protoplasts are incubated in the presence of Ca^{2+} and Ca^{2+} ionophores. It is thought that phytochrome may induce the opening of Ca^{2+} channels in the plasma membrane (Kuhlemeier 1992 and references therein). The involvement of Ca^{2+} in signal transduction pathways has also been suggested by studies on ethylene-mediated responses (Raz & Fluhr 1992). Evidence implicating

protein kinases, Ca^{2+} and calmodulin, G-proteins and phosphoinositides in the signal transduction pathway has been obtained from non-plant systems (see Kuhlemeier 1992 and references therein). An early signal in flower development, florigen appears to be produced by leaves and transported to the vegetative shoot apex where it is thought to initiate floral transition (Bernier 1988). However, little is known of the mechanism of action of this, or other, intermediates in the signal transduction pathway in plants.

1.3.3.3. *Cis*-acting motifs required for RNA processing

Transcription of plant genes by RNA polymerase II is generally from a single initiation site, usually an adenosine located 32 $\pm$ 7 nucleotides downstream of the TATA box. Production of a functional mRNA requires complex post-transcriptional modifications including 5'-capping, 3' end-processing and intron splicing. The 5' cap structure of eukaryotic mRNAs is composed of a guanine nucleotide resulting from addition of GMP by guanyltransferase. The guanine is connected by an unusual 5'-5' triphosphate linkage and also bears methyl groups ($m^7G(5')pp(5')N$). This modification is present on all mRNAs transcribed and appears to only require the recognition of a free 5'-phosphate. As well as promoting initiation complex formation with the 43S ribosome initiation complex (Pain 1986), the cap is thought to serve as a stabilising element on mRNA in both the nucleus and the cytoplasm.

1.3.3.3.1. Polyadenylation

Processing of mRNA 3'-ends and RNA splicing are activities requiring the recognition of specific mRNA structures. A polyadenylate tail added in a post-transcriptional process, is found at the 3' end of almost all eukaryotic mRNAs, and may function in translation and the control of RNA stability (Jackson & Standardt 1990). There are indications from animal systems that polyadenylation is necessary for termination of transcription (Connelly & Manley 1988). Endonucleolytic cleavage of a primary mRNA transcript occurs at a particular site and is followed by poly(A) addition to the upstream RNA fragment (Wahle 1992). The downstream RNA fragment is degraded. In eukaryotic cells, two *cis*-acting sequence elements have been identified as being essential for accurate and efficient 3' end formation; the highly conserved AAUAAA located 10-30 nucleotides upstream from the poly(A) addition site (Montell *et al.* 1983, Wickens & Stephenson 1984) and the less conserved (G)T-rich sequences immediately 3' to the cleavage site (McLauchlan *et*

al. 1985). Compilation analysis has shown that an AAUAAA-like sequence is conserved at the 3' end of nuclear plant genes (Joshi 1987b) but a significant proportion of plant genes do not contain a perfect AAUAAA-signal, and deviant signals most often contain a different base in position six (Ingelbrecht *et al.* 1989). AAUAAA elements are not absolutely required for polyadenylation of mRNAs in plants, and are absent from some plant genes (Mogen *et al.* 1990).

The (G)T-rich sequence motif present downstream from the polyadenylation site may also be required for 3' end processing (Ingelbrecht *et al.* 1989). This sequence apparently does not occur at a fixed position relative to the polyadenylation signal (Joshi 1987b). A majority of mammalian genes contain a YGTGTTYY sequence approximately 30 bp downstream from the polyadenylation signal and disruption of this spatial relationship reduces the efficiency of 3' processing (Ingelbrecht *et al.* 1989). In the 3' region of the octopine synthase gene (*ocs*) transcript, the YGTGTTYY-like motif is situated some 140 bp downstream from the major polyadenylation signal, whereas for the potato proteinase inhibitor gene, a similar sequence situated 46 bp downstream of the polyadenylation signal has been shown to be important for 3'-end formation (An *et al.* 1989). It has been demonstrated that chimeric genes containing only a polyadenylation signal at the 3' end are not efficiently expressed (Ingelbrecht *et al.* 1989).

Expression of deletion mutants has been used to show regions between 57 and 181 bases upstream from the CaMV polyadenylation site, and those between 60 and 137 bases upstream of the poly(A) addition site in pea *rbcS*, are required for efficient polyadenylation (Mogen *et al.* 1990). These studies have led to the hypothesis that interactions of different components of the plant polyadenylation apparatus with specific sequence elements and with each other are needed for efficient mRNA 3' end formation (Mogen *et al.* 1990).

Multiple polyadenylation sites are frequently found in mRNA transcripts from plant genes. These are usually preceded by multiple putative polyadenylation signals which are not necessarily conserved, even within a multigene family (Dean *et al.* 1986). For example, five possible polyadenylation signals are present within the patatin gene PS20 (Mignery *et al.* 1988), and Koes *et al.* (1986) found three potential polyadenylation signals and nine different poly(A) tail start sites among a number of cDNA clones encoding petunia CHS. Ingelbrecht *et al.* (1989) postulated that alternative usage of different polyadenylation signals could be caused by differences in the polyadenylation efficiencies of the motifs or by variation in the signal context. However, the multiple polyadenylation sites within the *ocs* 3' region have been found to be controlled by a series of far and near upstream elements as

well as by downstream elements (MacDonald *et al.* 1991). Mogen *et al.* (1992) have also demonstrated that specific, nonoverlapping, near and far upstream elements within the 3' region are required for efficient polyadenylation to occur at the various polyadenylation sites of a pea *rbcS* gene.

Post-transcriptional processes associated with 3'-end formation in plant mRNAs appear to be important in determining transcript prevalence (Ingelbrecht *et al.* 1989). This may be due to differential mRNA stability (Dean *et al.* 1985), or differential processing or transport of the transcripts (Ingelbrecht *et al.* 1989).

1.3.3.3.2. Transcript splicing

Introns (intervening sequences) are transcribed regions of genes that are not represented in mature mRNA, but are usually excised from primary transcripts (pre-mRNAs) by splicing within the nucleus (Sharp 1987). Introns are mostly localized in amino acid coding regions (Breathnach & Chambon 1981), however they have also been found in the 5'-untranslated regions of some plant genes, such as the actin genes from several plants (McElroy *et al.* 1990b) and the maize polyubiquitin genes (Christensen *et al.* 1992).

Removal of introns proceeds by a two-step mechanism. In the first step the pre-mRNA is cleaved at the 5' splice site and the 5'-terminal guanosine residue of the intron is linked to a specific adenosine residue (the branch point) upstream of the 3' splice site. In the second stage, cleavage at the 3' splice site and ligation of the exons occurs, producing accurately spliced RNA and the excised intron.

Conservation of various features of plant introns have been proposed, including a requirement for AU-rich sequences (Goodall & Filipowicz 1989) and a length of at least 70 nucleotides (Goodall & Filipowicz 1990). Although no plant intron branch site has been clearly characterized, lariat formation is assumed since it appears to be a universal feature of mRNA splicing, and it is supposed that plants have similar criteria for branch site selection as animals. Some evidence for this is that sequences matching the loose branch site consensus of animal introns CYRAY[5] can often be found upstream of the 3' splice site of plant introns (Brown 1986).

Differences can be shown in the purine or pyrimidine bias of intron splice junctions from monocot and dicot plants, with monocot junctions being predominantly pyrimidine-rich in the 3' splice junction, while dicot 3' splice sequences are higher in purine bases (Hanley & Schuler 1988). Studies of splicing

[5]where R is a purine (G or A)

in maize protoplasts indicate that monocots may not have the absolute requirement for A+U-rich sequences shown by dicot plant introns (Goodall & Filipowicz 1991). The dinucleotide frequencies of monocot and dicot genes differ in both exons and introns, suggesting that there may be differences in the mechanisms of splice-site recognition (White *et al.* 1992). Consensus sequences for splice sites in monocots and dicots have been proposed (eg. Hanley & Schuler 1988). White *et al.* (1992) have calculated a consensus sequence of AG | GUAAGU for plant 5' splice sites. Splice sites at the 3' end of introns were found by these authors to differ, with dicots preferring UGYAG | GG and monocots most commonly using UGCAG | GG. However, the only absolutely conserved sequences found in all plant introns are a GU dinucleotide at the 5' splice site and an AG dinucleotide at the 3' splice site (Goodall & Filipowicz 1989). It has been hypothesized that the splicing system achieves accurate recognition of intron borders by requiring the presence and appropriate positioning of several different sequence elements. These *cis*-acting elements are of variable importance in different organisms and various small ribonucleoprotein (snRNP) particles and proteins must bind in order to form a competent splicing machinery (Goodall & Filipowicz 1989).

The functional significance of introns is unknown, but recent studies have shown that in graminaceous monocots the presence of introns can enhance gene expression. For example, introns dramatically increased expression of the *adh1* gene and that of a chimeric CAT construct in maize, particularly when the intron was located near the 5' end of the mRNA (Callis *et al.* 1987). The presence of an intron within the 5' leader sequence of the rice *act1* transcript was essential for efficient gene expression from this promoter, and this activity was associated with a requirement for efficient intron splicing (McElroy *et al.* 1990c). Using the firefly luciferase and GUS reporter genes with the *adh1-S* intron 1 and the third intron of the maize actin gene, Luehrson & Walbot (1991) have demonstrated an enhancement of expression in transient assays at both the RNA and protein levels, together with an accumulation of unspliced transcripts. The enhancement effect was thought to be associated with intron splicing and not with the intron acting as a transcriptional enhancer. Higher levels of pre-mRNAs of the *bronze-2* gene of maize accumulated in husk tissue of field-grown, as compared to glasshouse-grown, maize (Nash *et al.* 1990), and an increased level of polyubiquitin transcripts containing an unspliced 5' untranslated-region intron was found in heat shocked maize seedlings in association with higher levels of transcription (Christensen *et al.* 1992). It appears possible that stress, in particular heat stress, can affect expression of some monocot genes by increasing accumulation of unspliced mRNA.

1.3.3.4. Translation

Although translational regulation of plant gene expression has not been studied to the same degree as transcriptional gene regulation, the expression of some proteins has been found to be affected at the translational level. After amaranth seedlings were transfered from light to dark, mRNA levels for both *rbcS* and the chloroplast-encoded *rbcL* remain unchanged for at least six hours, but incorporation of ^{35}S-methionine into the encoded proteins ceased completely within two hours (Berry *et al.* 1990). Furthermore, differential translation rates of heat shock mRNAs were found in callus and globular embryos of carrot, and it was suggested that this could be mediated by specific recognition elements present in the 5' untranslated regions, the length of the poly(A) tail, or modification of the translational apparatus (Apuya & Zimmerman 1992).

1.3.3.4.1. Translation initiation signals

Initiation usually occurs at the most 5' AUG codon of mature eukaryotic RNA. This is the basis of the scanning hypothesis wherein binding of the small ribosomal subunit occurs at the capped 5' end and is followed by the ribosome advancing along the RNA until an AUG sequence is encountered (Kozak 1981). At this site, the 80S ribosome is assembled and peptide synthesis begins. Translation initiation requires a variety of initiation factors, some of which are thought to bind to 40S ribosomal subunits prior to the association of the ribosomal subunit with the mRNA.

The initial interaction in translation is between the mRNA cap and a 25 kDa cap recognition protein (p25, also known as eIF-4E or CBP I). This protein is, or becomes, associated with two other subunits, eIF-4A (p44), a putative RNA helicase, and a 220 kDa protein (p220) of unknown function, to form eIF-4F. The next step is helix unwinding and then a further factor, eIF-4B triggers the expulsion of the eIF-4E and p220 subunits from the eIF-4F complex bound to the mRNA. These released components can then recruit free eIF-4A to reconstitute active eIF-4F. The eIF-4A subunit and eIF-4B remain bound to the mRNA, and can participate in the further unwinding of mRNA secondary and tertiary structure. At some point unknown in the eIF-4F pathway, the 43S ribosomal complex binds and proceeds to scan the mRNA for an AUG codon (see Thach 1992 for a review).

Examination of 79 plant genes concluded that 92% of these used the first AUG codon to initiate translation (Joshi 1987a). Translation of RNA into protein sequences in eukaryotes initiates at high efficiency at a consensus sequence of CCRCCAUGG, known as a Kozak recognition site (Kozak 1981). If the nucleotides

flanking a methionine codon are optimal, virtually all ribosomal initiation complexes are thought to stop and initiate at that AUG (Kozak 1981). Ribosomes are thought to be capable of alternative initiation at a downstream site provided that all of the upstream AUG codons are flanked by unfavourable sequences, this is known as the 'leaky scanning' model (Kozak 1986). Multiple in-frame ATG codons are frequently found at the 5' of plant mRNAs, for example the cDNA clone of a barley gene expressed in aleurone layers showed four potential translation initiation sites (Klemsdal *et al.* 1991).

Two independent computational analyses have found that initiation of translation in plant genes may prefer a consensus sequence differing from the Kozak recognition site. Joshi (1987a) reported a consensus of UAAACAAUGGC, while Lütcke *et al.* (1987) found a consensus of AACAAUGGC. A more recent analysis showed that monocot ribosomes preferred a translation initiation sequence of GCARCCAUGGC while dicot initiation sites tended to favour AAAAAMAUGGC[6], with the bases in positions -3 and +4[7] being most strongly conserved (Cavener & Ray 1991). Derivations of consensus sequences are heavily reliant on the composition of the sequence database, which tends to be skewed by the predominance of highly expressed genes and genes from species which are principally studied. It has been suggested that the Kozak recognition site is in fact that which provides exceptionally high levels of translation and, therefore, represents an optimal rather than a consensus sequence (Cavener & Ray 1991). The presence of non-consensus regions flanking translation initiation motifs may indicate these sequences play a role in determining the efficiency of translational initiation. A more accurate analysis of the influence of flanking sequences on translation initiation in plants awaits a more representative sample of plant genes becoming available.

1.3.3.4.2. Translation termination signals

Termination of translation is directed by one of three possible stop codons (UAA, UAG or UGA) and leads to the release of the completed polypeptide from the ribosome (Tate & Caskey 1990). The recognition of stop codons is thought to be mediated through specific proteins known as release factors. A eukaryotic release factor has been found in mammals and insects, which is believed to possess a recognition site for the three stop codons (Brown *et al.* 1990). However detailed

[6]where M is either A or C

[7]where the A of AUG is +1

information is still lacking on the method of action and the structure of release factors.

Misreading or bypassing of a stop codon has been well documented (see Hatfield *et al.* 1990), but for the majority of cellular mRNAs translation terminates efficiently at the stop codon. It has been suggested that the nucleotide in the fourth position may also influence the efficiency of the stop signal, and a survey of published sequences concluded that the signal UAA(A/G) was favoured in dicots, while UGA(A/G) was preferred in monocots. A bias toward the presence of G in the second or fourth positions could be related to a high G+C content of the mRNA (Brown *et al.* 1990). This tendency was also shown with data collated by Cavener & Ray (1991), who demonstrated a highly significant regression between G+C content and the presence of G in the translation termination signal. This larger data set supported a purine bias at the position immediately downstream of the stop codon, although there was also an appreciable quantity of U. These same tetranucleotide preferences were also observed by Angenon *et al.* (1990).

Termination of protein synthesis may play an important role in the regulation of expression. The study of Brown *et al.* (1990) indicated that within gene families, highly expressed genes retained the use of a particular stop codon regardless of variation in G+C composition, suggesting that powerful selection for a subgroup of signals is correlated with efficiency of gene expression. Therefore termination may prove to be an important mechanism in protein synthesis and overall expression.

1.4. Aims and strategy

1. To purify and sequence a clone encoding the alliinase subunit.

2. To analyse the structure of this clone and that of the deduced alliinase protein sequence.

3. To initiate studies into the representation of alliinase-encoding genes in the *A. cepa* genome and the expression of alliinase in *A. cepa* tissues.

The rapid development of gene cloning and plant transformation techniques has enabled the successful modification at the molecular level of several commercially important crop plant species that have formerly been recalcitrant to manipulation by traditional breeding strategies. As alliinase controls a unique and central step in flavour release, it was envisioned that the molecular cloning of the *A. cepa* alliinase gene would not only lead to a better understanding of *A. cepa* flavour release, but would also provide insights into the expression of this highly abundant *Allium* enzyme, and be an initial step toward developing molecular strategies ultimately aimed at changing the flavour of agronomically valuable *Allium* species.

This project was designed to complement and augment a parallel investigation aimed at isolating an alliinase clone which had been initiated by J.E. Lancaster of Crop and Food Research Ltd., Lincoln. At the time the work described in this thesis had begun, construction and screening of an *A. cepa* root tip λgt10 cDNA library with redundant oligonucleotides designed by back-translation of alliinase peptide sequences was in progress. As it was expected that alliinase cDNA clones which could be used as molecular probes would in due course become available, the present study initially focused on the construction of an *A. cepa* genomic library. It was anticipated that genomic clones would facilitate the identification and analysis of promoter regions involved in the regulation of alliinase gene expression.

Chapter 2. Materials and methods

2.1. Alliinase purification and peptide sequencing

Alliinase protein from the *Allium cepa* cultivar Pukekohe Longkeeper was purified by a protocol based on the method of Tobkin & Mazelis (1979) by M.L. Shaw (Crop and Food Research Ltd., Lincoln). The final purification step was size-exclusion chromatography using a Sepharose CL-6B gel filtration column (Clark *et al.* 1993). Alliinase had been purified 12-fold to a specific activity of 1625 nkat/mg as determined by the method of Schwimmer & Mazelis (1963). A single band was observed after Laemmli SDS-PAGE and silver staining.

N-terminal alliinase peptide sequencing was performed by Dr A. Carne, Department of Biochemistry, University of Otago. This sequence was confirmed, and further sequences were determined from alliinase peptides generated by cyanogen bromide fragmentation by Dr D.L. Christie, Department of Biochemistry, School of Biological Sciences, University of Auckland.

2.2. Genomic library construction and screening with synthetic oligonucleotides

2.2.1. Construction of an *A. cepa* genomic library

2.2.1.1. Insert preparation

2.2.1.1.1.DNA isolation

A. cepa DNA was isolated following the protocol previously described for bulb tissue by Kirk *et al.* (1970). *A. cepa* (cultivar Pukekohe Longkeeper) bulb tissue (200 g) was crushed in extraction buffer (40 ml of 0.5 M sucrose, 0.05 M Tris-HCl, 0.1 M EDTA pH 7.2) and filtered through 2 layers of muslin cloth. The extract was centrifuged at 0°C at 1 000 g for 15 minutes and the resulting pellet resuspended in 10 ml of extraction buffer with 330 μl of TritonX-100. This nuclear suspension was centrifuged at 1 000 g for 15 minutes. The resuspension step in extraction buffer with TritonX-100 was repeated twice, and the isolated nuclei were then lysed in 10

ml of buffer containing 0.1 M Tris-HCl, 0.1 M EDTA (pH 8.4), 0.5% (w/v) SDS and 500 μg/ml Proteinase K, and incubated at 37°C for 60 minutes. The lysate solution was adjusted to 1 M sodium perchlorate, and extracted three times with chloroform:isoamyl alcohol (24:1). The aqueous phase was then ethanol precipitated (Subsection 2.8.5). Precipitated DNA was spooled out with a glass pasteur pipette and dried under vacuum at room temperature. The dried DNA pellet was dissolved in $T_{10}E_1$ (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

2.2.1.1.2. Fragmentation of *A. cepa* DNA with *Sau* 3A

A. cepa DNA was partially digested with the restriction endonuclease *Sau* 3A to generate fragments with a predominant size range of 16 to 25 kb. The degree of digestion was controlled by stopping the reactions at an appropriate time point previously determined by carrying out 30 μl trial digestions containing DNA at concentrations between 0.12 μg/μl to 0.5 μg/μl, and *Sau* 3A at concentrations between 0.01 units/μl to 3 units/μl. DNA was dissolved in 1x *Sau* 3A digestion buffer (Gibco BRL) and *Sau* 3A added. Digestions were halted by the addition of EDTA (pH 8.0) to a concentration of 33 mM. The degree of digestion was determined by separating digestion products (usually in a volume containing 1 μg of DNA) on 0.8% agarose gels and comparing the position of the resultant DNA smear to commercial DNA size standards (see Subsection 2.8.1). Trial digestions gave differing rates of digestion for different DNA preparations and batches of *Sau* 3A. When a particular ratio resulted in the required degree of partial digestion, this was maintained in all subsequent reactions. Multiple 30 μl reactions were used to produce sufficient appropriately fragmented DNA for size fractionation, as scaling up these reactions was found to alter the degree of digestion. Digested DNA to be used for size fractionation was ethanol precipitated (described in Subsection 2.8.5) and resuspended in $T_{10}E_1$ at a concentration estimated from the amount of DNA in the initial reaction to be approximately 1 μg/μl.

2.2.1.1.3. Size fractionation of *A. cepa* DNA by velocity ultracentrifugation

Sau 3A-partially digested *A. cepa* DNA was size fractionated on 12 ml linear NaCl gradients (5-25% (w/v) NaCl in 3 mM EDTA pH 8.0) (Kaiser & Murray 1986). Digested *A. cepa* DNA (100 μg) was layered on the top of each gradient, which were then centrifuged for 4 hours at 37 000 rpm in an SW41 Beckman rotor at 12°C. Fractions (1 ml) were collected from a needle hole puncture in the base

of the centrifuge tube and 8 μl samples were run on 0.8% agarose gels in 1x TBE (8.9 mM Tris base, 8.9 mM boric acid, 1.0 mM EDTA pH 8.0) alongside commercial DNA size standards (see Subsection 2.8.1). Fractions containing DNA fragments of the required size, ie. between 16 and 25 kb were then pooled and ethanol precipitated as described in Subsection 2.8.5. DNA pellets were dried in a vacuum desiccator and resuspended in a small volume of $T_{10}E_1$.

2.2.1.2. Ligation of *A. cepa* DNA to λ-EMBL3 vector arms

The vector λEMBL3 was purchased pre-digested with *Bam* HI (Promega). Two parameters are of major importance when setting up ligation reactions. Firstly, the total DNA concentration in the ligation mixture must be high enough to ensure that intermolecular ligations, which lead to concatamers, are favoured over self-ligation, which reduces the yield of viable recombinant λ DNA molecules. Secondly, the ligation reaction should contain equimolar concentrations of the cloning site ends of each of the three species of DNA (left arm:insert:right arm) or a molar ratio of 2:1 for annealed λ arms to target insert DNA fragments. Errors in estimating DNA concentrations and damaged cloning sites can produce *in vitro* efficiencies lower than expected, therefore a series of pilot reactions were carried out to empirically determine the optimum ratio of λ arms to target insert DNA fragments.

To determine the optimal ligation conditions, 5 μl trial ligations were performed with insert:arms molar ratios of 1:9, 1:4, 1:1.5, and 1:0.9, assuming an average insert size of 15 kb. These trial ligation reactions contained 20 ng (1:9), 40 ng (1:4), 120 ng (1:1.5) and 200 ng (1:0.9) of size fractionated *A. cepa* DNA (100 ng = 0.01 pmole) with 500 ng of EMBL3-*Bam* HI arms (500 ng = 0.0175 pmole) in ligation buffer (40 mM Tris-HCl pH 7.4, 12.5 mM $MgCl_2$, 12.5 mM DTE, 1.25 mM ATP and 62.5 μg/ml BSA) (Sambrook *et al.* 1989) with 1-2 units of T4 DNA ligase and either 1 mM hexamine cobalt chloride (HCC) (Rusche & Howard-Flanders 1985) or 15% (w/v) PEG 6000 (Zimmerman & Harrison 1985). These conditions were calculated to give optimum concentrations of DNA termini (Dugaiczyk *et al.* 1975). Ligation reactions were incubated overnight at 4°C. Control ligations were identical but contained a test insert pTI11 (Promega), no ligase, or no insert DNA.

To ensure only concatenated DNA was packaged, 0.5 μl of the ligation reactions were examined for an increase in fragment size on 0.3% agarose gels (Subsection 2.8.1). The successful ligations were then packaged using a commercial packaging system (Promega Packagene) and titred by plating a series of dilutions in SM buffer

(20 mM Tris-HCl pH 7.4, 100 mM NaCl, 10 mM $MgCl_2$, 0.1% (w/v) gelatin) on the *E. coli* strain LE392 prepared as described in Subsection 2.8.6. Aliquots of diluted bacteriophage solutions (usually 20-50 μl) were combined with 200 μl of prepared *E. coli* LE392 and incubated at 37°C for 15 minutes. The infected cells were added to 4 ml of molten LB top agar (LB (1% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl) with 0.7% agar) containing 10 mM $MgSO_4$, and poured onto LB plates. Plaques were grown overnight at 37°C. Further ligations using the optimal conditions determined by the trial ligations were carried out and these were also packaged into bacteriophage and titred on *E. coli* LE392.

2.2.2. Design of redundant oligonucleotides for homology probing

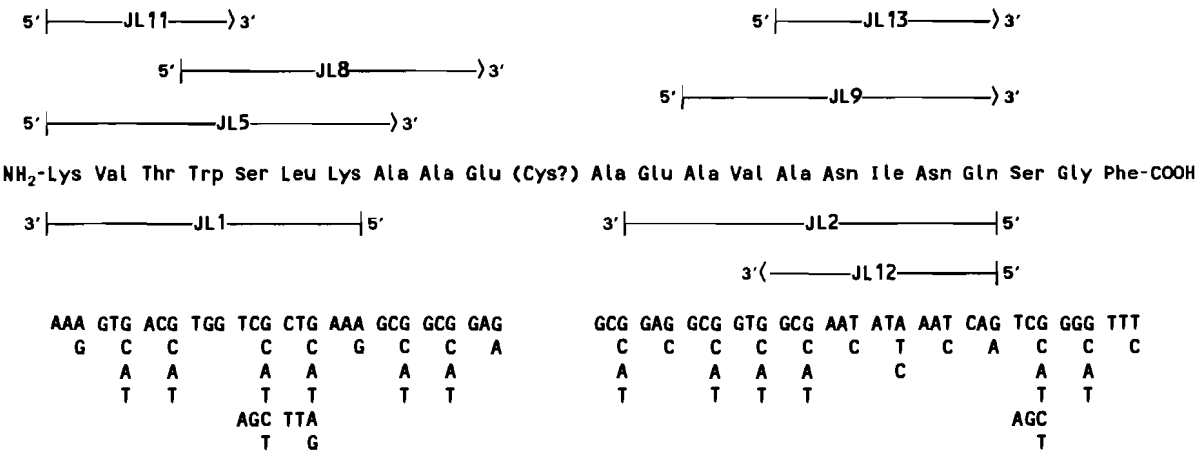
At the time this project was initiated, only the N-terminal 23 residue sequence of alliinase (ALN1) was known. The probes JL1 and JL2 were designed for homology probing by back-translation of this amino acid sequence (see Figure 2.1) and were made as long as possible to optimize hybridization specificity. The two oligonucleotides covered 15 of the 23 amino acid residues. The residue at position 11 was ambiguous, and had been tentatively identified as a cysteine. Also, the three residues at the C-terminus were considered to be less reliably determined and so were avoided in the design of probes. Back-translation of the available amino acid sequence resulted in highly redundant oligonucleotides, therefore a codon usage analysis of monocot gene sequences (Murray *et al.* 1989) was consulted to select preferred codon choices and thereby lower probe complexity.

The JL1 and JL2 oligonucleotides (shown in Table 2.1), were manufactured by the Molecular Pathology Laboratory at the Christchurch School of Medicine and supplied as a powder. Oligonucleotides were made up to a concentration of 200 μM in deionized water.

Figure 2.1. Alignment of redundant oligonucleotides designed by back-translation to alliinase peptide sequences

Oligonucleotides were designed from the N-terminal 23 amino acids of alliinase (ALN1) and a further 14 amino acid sequence of an alliinase fragment generated by cyanogen bromide cleavage (ALN3). The identity of the residue at position 11 of ALN1 (Cys?) was ambiguous. Oligonucleotides represented above the amino acid sequence were designed to the sense strand, oligonucleotides represented below the amino acid sequence were designed to the antisense strand. Probes with arrowheads were designed for PCR amplifications, while probes with flat ends were designed for homology probing. The DNA sequence represents all possible codon choices for the sense strand.

N-terminal amino acid sequence, ALN1



Cyanogen bromide-generated internal amino acid sequence, ALN3

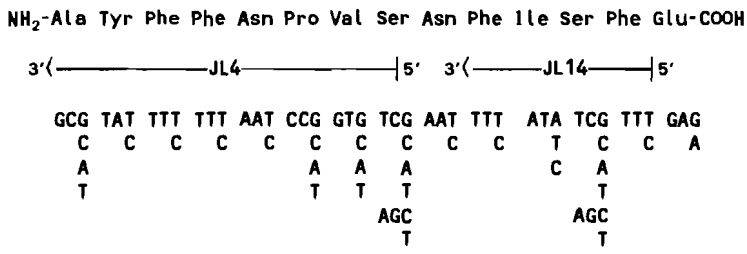


Table 2.1. Redundant oligonucleotides designed for homology probing and PCR by back-translation of alliinase peptide sequences

Oligonucleotide Sequence	Length	-fold redundancy
JL1 5′ CTT GAG GGA CCA GGT GAC CTT 3′ C CCT C C T A A T	21	384
JL2 5′ CTG GTT GAT GTT GGC GAC GGC CTC 3′ T A C C C A A	24	84
JL4 5′ ⊢SalI⊣ GAG ACG GGG TTG AAG AAG TAC GC 3′ CTC C A A A G A A A T T	23	1536
JL5 5′ ⊢SalI⊣ AAA GTG ACG TGG TCG CTG AAA GC 3′ G C C C C G A A A A T T T T	23	1024
JL8 5′ TGG TCG CTG AAA GCG GCG GA 3′ AGC T C G C C T T T T A A A A	20	4096
JL9 5′ GCG GTG GCG AAT ATA AAT CA 3′ C C C C T C T T T C A A A	20	768
JL11 5′ ⊢SmaI/BamHI⊣ AAA GTA ACG TGG 3′ G G C C T T A	12	32
JL12 5′ ⊢SmaI/BamHI⊣CTG ATT TAT ATT GGC 3′ T G A G C G T A	15	96
JL13 5′ ⊢SmaI/BamHI⊣GCA AAT ATA AAT CA 3′ G C T C C C T	14	48
JL14 5′ ⊢SmaI/BamHI⊣GAA AGT TAT GAA 3′ A GCA G A C A T	12	192

⊢SmaI/BamHI⊣ = 5' CCCGGGGATCC 3', ⊢SalI⊣ = 5' GGTCGACC 3'.

2.2.3. Genomic library screening

2.2.3.1. Capillary transfer of genomic library DNA to membranes

The un-amplified *A. cepa*-λEMBL3 genomic library was plated with *E. coli* LE392 in LB top agarose (LB with 0.7% agarose) on 20 plates at a density of approximately 13 000 pfu/ 90 mm petri dish as described in Subsection 2.2.1.2. Plaques were grown for 5 hours at 37°C (Slightom & Drong 1988) until they were just visible, at which stage they were transfered to 4°C. Two Hybond C nitrocellulose filters (Amersham) were cut to size and placed onto the agar surface of each plate for 1 minute to allow capillary transfer of DNA. Needle-holes were made in three locations at the periphery of the filters to enable later realignment. Uninfected bacterial lawns were used for control plates. Filters were soaked in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 5 minutes, followed by 2 washes of 5 minutes in neutralizing solution (1.5 M NaCl, 0.5 M Tris-HCl pH 8.0) and 1 hour in 2x SSC (0.3 M NaCl, 30 mM sodium citrate) then air-dried and baked in a vacuum oven at 80°C for 90 minutes (Kaiser & Murray 1986).

2.2.3.2. End-labelling of oligonucleotide probes

End-labelling utilizes the ability of the enzyme T4 polynucleotide kinase to catalyse the transfer of the γ-phosphate of ATP to a free 5'OH terminus of double or single stranded DNA or RNA. This reaction is ideally suited to labelling small synthetic oligonucleotides which are not amenable to random priming.

Probes were end-labelled by diluting the redundant oligonucleotide to a concentration of 2 μM (ie. approximately 300 pmole ends) in 10 μl of kinase buffer (0.05 M Tris-HCl pH 7.6, 0.01 M $MgCl_2$, 5 mM DTE, 0.1 mM spermidine-HCl, 0.1 mM EDTA) (Sambrook *et al.* 1989) with 30 μCi [γ-^{32}P]-dCTP (1000 Ci/mmol) and 10 units of T4 polynucleotide kinase. The reaction was incubated at 37°C for 3 hours. To ensure that the [γ-^{32}P]-dCTP labelling reaction went to near-completion, approximately 1000 cpm from the reaction was spotted onto a PEI cellulose TLC plate which was developed in 0.75 M NaH_2PO_4 pH 3.5. The resulting plate was exposed to Kodak X-omat AR film overnight at -70°C. Unincorporated [γ-^{32}P]-dCTP migrates in this system, whereas labelled probe remains at the origin. The probes generated were added directly to the hybridization solution.

2.2.3.3. Oligonucleotide hybridization conditions

Appropriate hybridization temperatures for the oligonucleotide probes JL1 and JL2 in 6x SSC (0.9 M NaCl, 90 mM sodium citrate) were calculated by the equation;

$$T_m = 81.5 + 16.6\,(\log_{10}[Na^+]) + 0.41(\%G+C) - (600/N)$$

(Meinkoth& Wahl 1984),

where T_m is the temperature at which a perfectly matched hybrid will be ½ dissociated and N is the oligonucleotide length. Values were calculated assuming a perfect hybrid with 50% G+C content. This gave values of 73°C for JL1 and 77°C for JL2. The reaction temperature affects the rate of hybridization, with the rate increasing to reach a broad maximum which is 20 - 25°C below *T*m for DNA-DNA annealing (Anderson & Young 1988). Therefore the incubation temperature (*T*i) was calculated as;

$$Ti = \frac{Tm(JL1) + Tm(JL2)}{2} - 20$$

Filters were washed for 1 hour at 65°C in 5x SSC, 0.5% SDS, and were then pre-hybridized in 6x SSC, 5 x Denhardt's solution (1% (w/v) ficoll, 1% (w/v) polyvinylpyrolidine, 1%(w/v) BSA), 0.1% SDS and 100 μg/ml sheared herring sperm DNA for 5 hours at 65°C.

The labelled oligonucleotide probes JL1 and JL2 were added together to the filters in pre-hybridization buffer (2 ml per filter), and hybridization was performed overnight in heat sealable plastic bags with gentle agitation at 55°C. Filters were washed four times in 2x SSC + 0.1% SDS for 10 minutes, and once in 1x SSC + 0.1% SDS for 90 minutes. They were then blotted dry and exposed to Kodak X-omat AR film overnight at -70°C. Positive signals (ie. those that aligned on duplicate filters) were used to identify potential positive plaques on the original plates. These were isolated from the plate in a plug of agar, placed into 500 μl SM buffer with 20 μl chloroform and vortexed. These selected plaques were re-screened at low density (100 to 200 pfu per plate) with conditions identical to the first round of screening.

2.3. Polymerase chain reaction (PCR) methodology

2.3.1. Design of redundant oligonucleotides for PCR

Oligonucleotide primers for PCR containing between 12 and 23 bases were designed from alliinase amino acid sequence data. The location of oligonucleotides relative to the alliinase peptide sequence is shown in Figure 2.1. The sequences and redundancy of the primers are shown in Table 2.1. JL5 and JL4 incorporated a site at the 5' end for the restriction endonuclease *Sal* I to facilitate later cloning, and were partially limited in their redundancy by making codon usage decisions based on monocot codon usage analysis (Murray *et al.* 1989). Two other sets of primers were designed for PCRs - JL8 and JL9 were 20 nucleotides long, and had redundancies of 4096-fold and 768-fold respectively. JL11, JL12, JL13 and JL14 were shorter primers (between 11 and 14 bases) of lower redundancy and contained sites for the restriction endonucleases *Sma* I and *Bam* HI at their 5' termini. This was intended to facilitate later cloning of products. The sequence of the primers was designed to minimize complexity but encourage specificity by using regions of the amino acid sequence of lowest codon redundancy when possible. All PCR primers, except for JL4 and JL5, incorporated every possible coding option.

These oligonucleotides were manufactured by commercial suppliers (Oligos etc. Inc. Connecticut), except for JL5 and JL4 which were synthesized at the Molecular Pathology Laboratory, Christchurch School of Medicine. Oligonucleotides were supplied in powder form and were dissolved in sterile distilled water to a concentration of 200 μM. The oligonucleotide $(dT)_{17}$ (Promega) was also used as a non-specific primer in some amplifications.

2.3.2. PCR template preparation

2.3.2.1. *A. cepa* genomic DNA preparation

A. cepa DNA had shown significant resistance to cleavage by the restriction endonuclease *Sau* 3A in preceding studies. This may be due to the presence of enzyme-inhibiting polysaccharides or charged polymers which are known to co-purify with plant DNAs (Rogers & Bendich 1988, Slightom & Drong 1988). For this reason subsequent *A. cepa* DNA isolations were carried out by a modified

method of Lassner *et al.* (1989) which utilizes the differential solubilities of polysaccharides and nucleic acids in the presence of CTAB to separate these compounds.

CTAB buffer (0.14 M sorbitol, 0.22 M Tris-HCl pH 7.5, 22 mM EDTA, 0.8 M NaCl, 0.8% (v/v) CTAB and 1% N-lauryl sarcosine) was pre-heated to 60°C. Just prior to use β-mercaptoethanol was added to a final concentration of 0.2%. Inner leaves were dissected from sprouting *A. cepa* bulbs (cultivar Southport White Globe) and 0.5 g placed in a plastic zip-top bag with 6 volumes (v/w) of CTAB buffer. *A. cepa* tissue was thoroughly crushed into the buffer using a small wallpaper roller. The resulting viscous solution was used to fill 1.5 ml eppendorf tubes containing 300 μl chloroform, and mixed thoroughly by inversion. Extractions were incubated at 65°C for 30 minutes and then centrifuged at 12 000 g for 5 minutes to separate the phases. The aqueous phase was mixed with 600 μl isopropanol and precipitated DNA collected by centrifugation (12 000 g, 5 minutes) at room temperature. The DNA pellet was rinsed with 70% ethanol, dried under vacuum at room temperature and resuspended in $T_{10}E_1$.

2.3.2.2. Complementary DNA (cDNA) template preparation

Total and poly(A)+ *A. cepa* RNA extracted from root tip tissue of the cultivar Southport White Globe were kindly donated for this experiment by J. Farrant (Crop and Food Research Ltd., Lincoln). Preparation of a first strand cDNA template was carried out by the method of D'Alessio *et al.* (1987). *A. cepa* poly(A)+ RNA (50 ng) or 1 μg of total *A. cepa* RNA was mixed with 500 ng of $(dT)_{17}$ in a 7.5 μl reaction volume, heated to 70°C for 10 minutes and then chilled on ice. To this was added 4 μl of 5x reverse transcriptase (Superscript) reaction buffer (Gibco BRL), 2 μl of 0.1 M DTT, 0.5 μl (200 units) of RNasin (Promega), 5 μl of dNTP stock (2 mM dATP, dTTP, dCTP, dGTP) and the reaction was incubated at 37°C for 2 minutes. Moloney murine leukemia virus (Mo-MuLV) reverse transcriptase (1 μl, 200 units) was added and the reaction incubated for 1 hour at 37°C. The first-strand cDNA was prepared for PCR amplification by the method of Domec *et al.* (1990). At the end of the incubation period 6.6 μl of 1 N sodium hydroxide was added to the reaction mixture and incubated for 30 minutes at 46°C. The first-strand reaction mixture was extracted once with water-saturated phenol and twice with chloroform. The cDNA was ethanol precipitated (Subsection 2.8.5) and dried. The cDNA pellet was dissolved in 28 μl of deionized distilled water.

2.3.2.3. Template enrichment by Southern hybridization with oligonucleotides

Increased specificity of PCR amplification can be achieved by selection of DNA templates after size-fractionation of restriction enzyme-digested genomic DNA (Beck & Ho 1988). *Triticum aestivum* (wheat) DNA was a gift from T. Frew (Crop and Food Research Ltd., Lincoln). *A. cepa* DNA (30 μg) was digested in a total volume of 30 μl containing 1x buffer (supplied by the manufacturer) with 10-20 units of the restriction endonucleases *Eco* RV, *Eco* RI, and *Hin*d III for 5 hours at 37°C. Wheat DNA was digested with *Hin*d III and *Eco* RV. Digested DNA was fractionated on 0.9% agarose gels in 1x TBE at 5 volts/cm. Agarose gels were soaked in 0.25 M HCl for 15 minutes to depurinate the DNA, followed by neutralization in transfer buffer (0.4 N NaOH, 0.6 M NaCl) for 20 minutes. Transfer of DNA to positively charged nylon membranes (Zetaprobe, Biorad) was achieved by alkaline capillary transfer for four hours (Reed & Mann 1985). Nylon membranes were rinsed in 0.2 M Tris-HCl pH 7.5, 2 x SSC and air-dried.

Membranes were washed in 2x SSC for 15 minutes at 42°C then transferred to pre-hybridization buffer (6x SSC, 10% (w/v) dextran sulphate, 1% SDS, 0.1% (w/v) sodium pyrophosphate and 0.3 mg/ml sheared herring sperm DNA) at 42°C for 2 hours. Hybridizations were performed by end-labelling the oligonucleotides JL4, JL5, JL8 and JL9 with [γ-^{32}P]-dCTP as described for genomic library screening (Subsection 2.2.3.2), and adding these to replicate membranes in pre-hybridization buffer. Membranes were hybridized overnight at 42°C and washed the next day for 30 minutes each in 4x SSC + 0.1% SDS, 2x SSC + 0.1% SDS, 1x SSC + 0.1% SDS and 0.5x SSC + 0.1% SDS. Membranes were visualized by exposure to Kodak X-omat AR film overnight at -70°C with intensifying screens.

The size of bands providing a positive signal were calculated from the relative migration of 1 kb DNA molecular size standards (Gibco BRL). A duplicate gel of digested *A. cepa* and wheat DNA was stained by immersion in 0.5 μg/ml ethidium bromide and fragments of the sizes corresponding to positive signals were excised from the gel on a u.v. transilluminator. Fragments of equivalent size were also excised from control wheat DNA, although no hybridization of the oligonucleotide to this DNA was apparent. The DNA was removed from the agarose slices by centrifugation through silanised glass wool (Heery *et al.* 1990). Purified fragments were quantitated by comparing their fluorescence after ethidium bromide staining to that of DNA standards made from λ-*Hin*d III digested DNA (Gibco BRL) of known concentration.

2.3.3. Standard PCR reaction conditions

PCR was carried out with the templates described above. The conditions used are summarised in Table 2.2 on page 64. To ensure more than 10^4 copies of the alliinase gene were present (assuming a single copy gene), 50 ng of intact *A. cepa* DNA was used. Otherwise 25 ng of the enriched templates selected by Southern hybridization, or 5 μl from the resuspended cDNA preparation were used. Nested PCRs were carried out in a final volume of 25 μl using a 5 μl aliquot of the first amplification. PCR was generally carried out in 25 μl reaction volumes containing 1 μM of the appropriate primer, 0.2 mM each dNTP in 1x PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin) and 0.4 units *Taq* DNA polymerase (Boehringer Mannheim). Approximately 50 μl of mineral oil was layered on top of each reaction to prevent evaporation.

The standard PCR sequence consisted of 1 cycle of 95°C for 6 minutes, followed by 30 cycles of 95°C for 1 minute, 37°C for 1 minute, and 72°C for 2 minutes followed by a 6°C soak cycle. Control reactions consisted of only one of the two primers, or no template to check for possible DNA contamination. Positive controls consisted of the plasmid pCGN587 with GT5 and GT6 primers (gifted by Dr. G. Timmerman, Crop and Food Research Ltd., Lincoln). Each reaction was replicated to determine reproducibility.

Aliquots (usually 10 - 15 μl) of the PCR amplifications were loaded onto 1.2% agarose, or 2% Nusieve (FMC Bioproducts) + 1% agarose gels in 1x TBE and products were visualized by agarose gel electrophoresis as described in Subsection 2.8.1.

2.3.4. Isolation and identification of a PCR fragment

2.3.4.1. Isolation of a PCR product for reamplification

A PCR product was excised in a slice of agarose from gels and eluted by centrifugation through silanised glass wool (Heery *et al.* 1990). This eluent was then used as a template in further PCR amplifications by adding 5 μl to a 25 μl reaction. PCR conditions were optimized for these reactions. The parameters varied were; primer concentration (0.2 μM, 0.5 μM and 1.0 μM); magnesium ion concentration (1.5 mM, 2.0 mM, 2.5 mM or 3.0 mM); cycle numbers (18, 23, 26 or 30 cycles); annealing temperature (37°C, 40°C, 45°C, 50°C, 60°C).

2.3.4.2. Fragment isolation for subcloning

A fragment of interest was isolated by excision from a 1% low melting point agarose gel in 1x TBE. DNA was removed from the agarose slice by freeze/thawing and collected in the aqueous phase. The concentration of DNA was determined by comparing the intensity of fluorescence in the presence of ethidium bromide to that of standards prepared from λ DNA of known concentration (Gibco BRL).

2.3.4.3. Ligation to pUC19 vector

The fragment was prepared for ligation to a pUC19 vector by phosphorylation with T4 polynucleotide kinase. This 10 μl reaction contained 40 ng of the PCR product in kinase buffer (50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 5 mM DTE, 10 ng/μl BSA), 1 mM ATP and 10 units T4 polynucleotide kinase and was incubated at 37°C for 30 minutes followed by heat inactivation of the enzyme at 65°C for 10 minutes. The phosphorylated fragment was blunt-end ligated to a de-phosphorylated pUC19 plasmid that had been prepared by digestion with *Sma* I. The 15 μl ligation reaction contained the phosphorylated PCR product (approximately 40 ng), 120 ng dephosphorylated, *Sma* I-linearized pUC19 vector, 1x buffer as for the previous phosphorylation reaction, 1 mM ATP, 10 μM HCC, and 2 units of T4 DNA ligase. Negative controls contained no insert DNA. The reaction was incubated at 16°C for 8 hours and then diluted 1:13 with $T_{10}E_1$.

2.3.4.4. Transformation of TG1 competent cells

The recombinant plasmid and the negative control were transformed into *E. coli* TG1 which had been rendered competent with $CaCl_2$ (Sambrook *et al.* 1989). To transform the cells, 5 μl of the diluted ligation reaction was mixed with 50 μl of a solution of competent TG1 and incubated on ice for 30 minutes. The cells were heat shocked for 45 seconds at 42°C and re-incubated on ice for 2 minutes, then added to 1 ml of 2x YT (1% (w/v) bactotryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl) and incubated at 37°C for 1 hour with agitation. Aliquots (100 μl) of this broth was then spread onto LB agar plates containing 100 μg/ml ampicillin and 50 μg/ml X-gal. Putative recombinants grew as white colonies in the presence of X-gal (Messing 1983). The presence of a recombinant plasmid containing the fragment of interest was confirmed by placing part of a selected colony into 100 μl of water,

boiling for 5 minutes, and using 1 μl of the resulting solution as a template for PCR as described in Subsection 2.3.3.

2.3.4.5. Plasmid DNA isolation

Plasmid DNA was isolated from positive recombinants by a mini-alkaline lysis method (Sambrook *et al.* 1989). Cells were grown overnight in 40 ml of LB supplemented with 100 μg/ml ampicillin and harvested the next morning by centrifugation at 7 000 g for 10 minutes at room temperature. The resulting cell pellet was resuspended in 1.6 ml of 50 mM glucose, 25 mM Tris-HCl pH 7.5, 1 mM EDTA. This solution was adjusted to a concentration of 1.25 mg/ml lysozyme and incubated at room temperature for 15 minutes. Cells were lysed by the addition of 3.2 ml of lysis solution (0.2 N NaOH, 1% SDS) followed by a 5 minute incubation at room temperature. Sodium acetate (2.4 ml of 3 M, pH 5.4) was then added and the solution gently mixed and incubated for 15 minutes on ice. The lysed cell suspension was centrifuged at 27 000 g for 15 minutes and the supernatant removed and precipitated with two volumes of ethanol. DNA was immediately pelleted by centrifugation at 12 000 g for 10 minutes (4°C). The pellet was dried under vacuum at room temperature and then neutralized by the addition of 4 ml of 0.1 M sodium acetate, 0.05 M Tris-HCl pH 8.0, followed by a 3 minute incubation on ice. DNA was precipitated with 9.2 ml of ethanol, collected by centrifugation at 12 000 g for 10 minutes and again dried under vacuum. The plasmid pellet was dissolved in 200 μl of 10 μg/ml RNase A, then extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) followed by an extraction with chloroform:isoamyl alcohol (24:1). The aqueous phase was then ethanol precipitated (see Subsection 2.8.5) and collected by centrifugation at 12 000 g for 10 minutes. After drying *in vacuo*, the DNA pellet was dissolved in $T_{10}E_1$ and its concentration determined spectrophotometrically.

2.3.4.6. Sequencing of a PCR product

Plasmid DNA was sequenced by the dideoxy-chain termination method (Sanger *et al.* 1977) with the Sequenase kit (United States Biochemicals) as described for double-stranded templates by the manufacturer. Sequencing reactions were primed with 17mer universal pUC/M13 forward and reverse primers (Promega).

Sequencing gels were cast between two glass plates with a 0.35 mm spacer. The smaller of the two plates was silanised by spreading 1-2 ml of silanising solution (2% dimethyldichlorosilane in trichloroethane) over the inner plate surface which

was then allowed to dry at room temperature. The larger plate was immersed in Bind-silane (0.4% γ-methacryloxypropyltrimethoxysilane, pH 3.5) for 1 hour. After assembly, approximately 60 ml of sequencing polyacrylamide (6% polyacrylamide, 8 M urea in 1x TBE, 0.05% ammonium persulphate, 0.05% TEMED) was poured between the plates to form the gel. Gels were allowed to polymerise for 1 hour. Immediately prior to loading, wells were formed using a 0.35 mm sharkstooth comb (Gibco BRL), and flushed out with 1x TBE buffer to remove free urea and unpolymerised acrylamide. Aliquots (3 μl) of each reaction were incubated at 80°C for 3 minutes and then loaded onto gels. Electrophoresis was in 1x TBE at 60 watts constant power for 2 hours (short gel) or 6 hours (long gel) using a Gibco BRL Model S2 sequencing gel electrophoresis apparatus. Glass plates were gently prised apart after electrophoresis, and gels were fixed by immersion in 5% glacial acetic acid and 15% methanol for 1 hour. Gels were dried on the glass plate at 80°C for 2 hours and visualized by exposure to Cronex film (Dupont) for up to 10 days.

Table 2.2. Summary of the polymerase chain reactions performed with redundant oligonucleotides designed from alliinase peptide sequences[1]

	Primers				Primer concentration (mM)		$MgCl_2$ concentration (mM)		Annealing temperature (°C)	
	First round		Second round		First round	Second round	First round	Second round	First round	Second round
Template	Sense	Antisense	Sense	Antisense						
1. cDNA										
	JL5	$(dT)_{17}$	-	-	0.5	-	1.5	-	55	-
			-	-	1.0	-	1.5	-	35,50	-
			-	-	10.0	-	1.5	-	35	-
			-	-	10.0	-	3.0	-	35,50	-
	JL5	JL4	-	-	1.0	-	1.5,3.0	-	35	-
	JL8	$(dT)_{17}$	-	-	1.0	-	1.5	-	50	-
			-	-	10.0	-	3.0	-	50	-
	JL8	JL4	-	-	1.0	-	1.5,30	-	35	-
	JL9	$(dT)_{17}$	-	-	1.0	-	1.5	-	50	-
			-	-	10.0	-	3.0	-	50	-
	JL5	$(dT)_{17}$	JL8,JL9	$(dT)_{17}$	2.0	4.0	1.5	1.5	45	45
			JL8,JL9	$(dT)_{17}$	4.0	4.0	1.5	1.5	45	45
	JL5	JL4	JL8	JL4	1.0	1.0	1.5	1.5	35	35
			JL8	JL4	1.0	1.0	3.0	3.0	35	35
	JL5	JL4	JL9	JL4	1.0	1.0	1.5	1.5	35	35
			JL9	JL4	1.0	1.0	3.0	3.0	35	35
	JL8	$(dT)_{17}$	JL9	$(dT)_{17}$	2.0	4.0	1.5	1.5	45	45
			JL9	$(dT)_{17}$	4.0	4.0	1.5	1.5	45	45
			JL9	JL4	4.0	4.0	1.5	1.5	45	45
	JL8	JL4	JL9	JL4	1.0	1.0	1.5	1.5	35	35

	Primers				Primer concentration (mM)		$MgCl_2$ concentration (mM)		Annealing temperature (°C)	
	First round		Second round		First round	Second round	First round	Second round	First round	Second round
Template	**Sense**	**Antisense**	**Sense**	**Antisense**						
			JL9	JL4	1.0	1.0	3.0	3.0	35	35
2. Genomic DNA										
	JL5	JL4	-	-	1.0	-	1.5	-	35	-
			-	-	1.0	-	1.5	-	40	-
			-	-	1.0	-	3.0	-	35	-
			-	-	1.0	-	3.0	-	40	-
	JL8	JL4	-	-	1.0	-	1.5	-	35	-
			-	-	1.0	-	1.5	-	40	-
			-	-	1.0	-	3.0	-	35	-
			-	-	1.0	-	3.0	-	40	-
	JL9	JL4	-	-	1.0	-	1.5	-	35	-
			-	-	1.0	-	1.5	-	40	-
			-	-	1.0	-	3.0	-	35	-
			-	-	1.0	-	3.0	-	40	-
	JL11	JL12	-	-	1.0	-	1.5	-	35	-
			-	-	1.0	-	1.5	-	50	-
			-	-	1.0	-	3.0	-	50	-
	JL11	JL14	-	-	1.0	-	1.5	-	50	-
			-	-	1.0	-	3.0	-	50	-
			-	-	1.0	-	1.5	-	35	-
	JL13	JL14	-	-	1.0	-	1.5	-	35	-
			-	-	1.0	-	1.5	-	50	-

	Primers				Primer concentration (mM)		$MgCl_2$ concentration (mM)		Annealing temperature (°C)	
	First round		Second round		First round	Second round	First round	Second round	First round	Second round
Template	**Sense**	**Antisense**	**Sense**	**Antisense**						
			-	-	1.0	-	3.0	-	50	-
	JL5	JL4	JL8	JL4	1.0	1.0	1.5	1.5	35	35
					1.0	1.0	3.0	3.0	35	35
	JL5	JL4	JL9	JL4	1.0	1.0	1.5	1.5	35	35
					1.0	1.0	3.0	3.0	35	35
	JL8	JL4	JL9	JL4	1.0	1.0	1.5	1.5	35	35
					1.0	1.0	3.0	3.0	35	35
3. Southern-selected DNA										
	JL8	JL4	-	-	1.0	-	1.5	-	37	-
	JL9	JL4	-	-	1.0	-	1.5	-	37	-
	JL11	JL12	-	-	1.0	-	1.5	-	37	-
			-	-	1.0	-	1.5	-	40	-
			-	-	1.0	-	1.5	-	50	-
	JL11	JL14	-	-	1.0	-	1.5	-	37	
			-	-	1.0	-	1.5	-	40	-
			-	-	1.0	-	1.5	-	50	-
	JL13	JL14	-	-	1.0	-	1.5	-	37	-
			-	-	1.0	-	1.5	-	40	-
			-	-	1.0	-	1.5	-	40	-
			-	-	1.0	-	1.5	-	50	-

[1] Standard reaction conditions are as outlined in Subsection 2.3.3.

2.4. Construction and immunoscreening of an *A. cepa* cDNA expression library

2.4.1. Polyclonal antibody production and titre

Anti-alliinase polyclonal antisera were produced in two New Zealand White rabbits (I65 and I66) by subcutaneous injection of 200 μg of alliinase in adjuvant. An identical booster shot was given 25 days after the first treatment, followed by a third identical treatment 15 days later. Polyclonal sera was collected and titred by ELISA against purified alliinase 51 days after the first treatment, and stored at -80°C. Antibody production and determination of reactivity were carried out by R. Hyder at the Surgical Pathology Laboratory, Christchurch School of Medicine.

2.4.2. Optimization of immunoscreening conditions

Dot blots were carried out to assess the ability of the alliinase antiserum to detect alliinase bound to nitrocellulose, and to determine the appropriate primary antibody dilution. These studies also were carried out to determine the avidity of the secondary antibody conjugate for the primary antibody, and to detect the presence of antibodies cross-reacting with *E. coli*/phage antigens. This approach followed that recommended by Stratagene for immunoscreening and is based on the western blotting protocol of Huynh *et al.* (1986).

Sixteen replicate nitrocellulose membranes were dot blotted with samples of alliinase at 100 ng, 10 ng, 1 ng, 100 pg and 10 pg, of I66 anti-alliinase polyclonal serum at dilutions of 1:500, 1:1 000, 1:10 000, 1:15 000 and 1:20 000 and an *E. coli*/bacteriophage lysate (Stratagene) at 1:1000, 1:100, 1:10 dilutions and undiluted. Membranes were immersed in blocking solution (1% BSA in 10 mM Tris-HCl pH 8.0, 150 mM NaCl (TBS)) overnight at 4°C and washed five times the next morning in TBS + 0.05% Tween-20 (TBST). A replicate dot blot was incubated in each combination of four dilutions of I66 anti-alliinase antiserum (1:5 000, 1:15 000, 1:30 000 or 1:60 000) and four dilutions of goat anti-rabbit alkaline phosphatase conjugate (GAR-AP) (1:300, 1:800, 1:1 200, 1:2 000). Membranes were incubated with I66 serum for 2 hours, and with GAR-AP for 1 hour, and washed 5 times for 5 minutes in TBST between each incubation. The membranes were finally washed 5 more times in TBST, and then immersed in 300 μg/ml NBT and 150 μg/ml BCIP in colour development solution (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$) for 30 minutes in the dark for colour development at sites of alkaline phosphatase activity.

2.4.3. Reactivity of anti-alliinase antibodies to deglycosylated alliinase

Polyclonal antibodies raised against plant glycoproteins can be predominantly reactive to the oligosaccharide side chain and show little or no reactivity to the protein backbone (Faye & Chrispeels 1988). Plant glycoproteins are expressed as aglycones in *E. coli*, as bacteria do not have the capacity to synthesize and attach oligosaccharide side chains. Consequently, confirmation of the affinity of polyclonal preparations to the protein backbone is important when they are to be used to screen cDNA libraries expressed in *E. coli*.

Determination of the reactivity of anti-alliinase antiserum to the alliinase protein backbone, and determination of the molecular mass of deglycosylated alliinase was carried out by M.L. Shaw. Alliinase was deglycosylated using trifluoromethane sulphonic acid (TFMS) following the protocol of Paaren *et al.* (1987). Deglycosylated and intact alliinase were separated by SDS-PAGE (Subsection 2.8.2) and transferred to nitrocellulose membranes by electroblotting (Subsection 2.8.4). The absence of oligosaccharide side-chains was confirmed by probing nitrocellulose membranes with concanavalin A-horseradish peroxidase conjugate (Clegg 1982) using a 4-chloro-1-napthol development procedure (De Blas & Cherwinski 1983). The affinity of the I66 anti-alliinase antisera to deglycosylated alliinase was determined by western blotting using conditions as described in Subsection 2.4.2.

2.4.4. cDNA library construction

2.4.4.1. RNA isolation

2.4.4.1.1. Extraction of *A. cepa* RNA

To minimize contamination of RNA samples with RNases, all glassware used for RNA work was baked overnight at 180°C and all plasticware used was new. All aqueous solutions (except those containing Tris-HCl) were incubated overnight at 37°C with 0.05% diethylpyrocarbonate (DEPC) and then autoclaved for 15 minutes at 121°C and 15 psi. Solutions that could not be incubated with DEPC were made up from DEPC-treated water and fresh reagents. Gel casting trays, boxes and combs were soaked in 3% hydrogen peroxide for 1 hour and rinsed thoroughly with RNase-free deionized water prior to use.

Trial extractions from *A. cepa* bulbs, basal plate tissue (containing the meristem), leaves and flower stems resulted in low RNA yields. RNA yields were highest from

A. cepa root tips and sprouting inner leaves from bulbs. As collection of root tip tissue was time consuming, RNA for cDNA library construction was extracted from sprouting bulb inner leaves. *A. cepa* bulbs (cultivar Southport White Globe) were sprouted in vermiculite in a glasshouse during early spring. This cultivar sprouted readily under glasshouse conditions and was used in preference to the brown-skinned cultivar, Pukekohe Longkeeper, which sprouted only occasionally. Once leaf tips could be seen emerging from the bulb, inner shoots were removed and used for RNA isolation. RNA extractions were based on the method of Logemann *et al.* (1987) for tissues with a high polysaccharide content. *A. cepa* tissue was cut into small pieces with a fresh scalpel blade, and ground in liquid nitrogen to a fine powder. Approximately 5 volumes (v/w) of guanidine-HCl buffer (8 M guanidine-HCl, 20 mM MES, 20 mM EDTA, 50 mM β-mercaptoethanol, pH 7.0) was added and grinding continued until tissue thawed to become a paste. To remove cell debris, the homogenate was centrifuged at 12 000 g for 10 minutes at room temperature. The supernatant was removed and extracted twice with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). The aqueous phase from this extraction was precipitated with 0.7 volumes of cold ethanol and 0.2 volumes of 1 M acetic acid at -70°C for 1 hour. After centrifugation at 12 000 g for 10 minutes at 0°C, the resulting pellets were washed 4-5 times with 3 M sodium acetate until all visible traces of gelatinous material had been removed. Pellets were collected between each wash by centrifugation at 12 000 g for 10 minutes at 0°C. The RNA pellet was then rinsed with 70% ethanol and dried under vacuum. The recovered RNA was resuspended in RNase-free deionized water. The concentration of RNA preparations was determined spectrophotometrically assuming a standard extinction value of 1.00 was equivalent to 40 μg/ml of RNA (Sambrook *et al.* 1989).

To check that no degradation of RNA occurred, aliquots from all samples were visualized by ethidium bromide staining after separation by electrophoresis. Dried samples of RNA (1-2 μg) were resuspended in 10 μl of loading buffer (48% (v/v) formamide, 6.4% (v/v) formaldehyde, 5.3% (v/v) glycerol and 5.3% (v/v) saturated bromophenol blue solution in MOPS buffer) and boiled for 2 minutes. Samples were separated on 1% agarose gels containing 2% formaldehyde in MOPS buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA), at 3 volts/cm overnight. RNA was stained by immersing the gel for 30 minutes in 0.5 μg/ml ethidium bromide, and destained for 3 hours in RNase-free water. Gels were examined as described in Subsection 2.8.1. The presence of defined 28S and 18S rRNA bands was taken as an indicator of sample integrity.

2.4.4.1.2. Extraction of poly(A)+ RNA

Poly(A)+ RNA was extracted from *A. cepa* total RNA using the PolyATtract extraction kit (Promega). This kit uses a biotinylated oligo(dT) primer that hybridizes to the 3' poly(A)+ region of mRNA species. Hybrids are captured and washed at high stringency using streptavidin coupled to paramagnetic particles which are trapped by a strong magnet against the wall of the reaction tube.

2.4.4.1.3. *In vitro* translation

The integrity of the poly(A)+ RNA was determined by the incorporation of ^{35}S-methionine into protein during *in vitro* translation reactions using a rabbit reticulocyte lysate kit (Gibco BRL). Controls consisted of a reaction lacking RNA (negative control), and a reaction containing haemoglobin RNA supplied with the kit (positive control).

The proportion of ^{35}S-methionine incorporated into proteins was determined by TCA precipitation of 5 μl aliquots of the 30 μl *in vitro* translation reactions. Duplicate aliquots were transferred onto separate glass-fibre filter papers (Whatman GF/C) before and after the reactions were incubated. Filters were washed four times for 5 minutes in an excess volume of cold 5% TCA, 10 mM sodium pyrophosphate. The filters were then rinsed with 70% ethanol and air-dried (Sambrook *et al.* 1989). Amounts of radioactivity on the filters were determined by liquid scintillation counting in a toluene-based scintillant (toluene/PPO/POPOP). Intact washed and unwashed filters were counted.

The protein products in 10 μl samples of the *in vitro* translation of *A. cepa* poly(A)+ RNA and the positive control reaction were fractionated on a 12% Laemmli SDS-PAGE gel as described in Subsection 2.8.2. The gel was fixed for 30 minutes in 50% (v/v) ethanol and 10% (v/v) acetic acid. The gel was dried and the radioactively labelled translation products were visualized by exposure to Kodak X-omat AR film for 14 days at -70°C.

2.4.4.2. cDNA synthesis

Complementary DNA (cDNA) was synthesized, and the cDNA library was created from *A. cepa* poly(A)+ RNA using the Stratagene ZAP-cDNA synthesis kit. This system employs a $(dT)_{18}$-*Xho* I linker-primer to initiate first strand cDNA synthesis catalysed by Mu-MLV reverse transcriptase. The resultant 3'-end location

of an *Xho* I restriction site allows the final double-stranded cDNA to be inserted into the λZAPII multifunctional phagemid vector in a sense orientation with respect to the *LacZ* promoter. dCTP is replaced by the analogue 5-methyl dCTP in the first strand reaction to prevent digestion of the cDNA by restriction enzymes used in subsequent cloning steps. The second strand was synthesized using RNase H to generate short oligonucleotide probes that function as primers for T4 DNA polymerase I. The cDNA ends were blunted with T4 DNA polymerase I and then ligated to linkers containing an *Eco* RI recognition sequence. Linkers were restricted with *Eco* RI and *Xho* I and removed by molecular sieving with Sephacryl S400.

The yields of cDNA synthesis reactions were determined using TCA precipitations to calculate the incorporation of a labelled nucleotide into DNA. Prior to the incubation, a 5 μl sample of the 50 μl first strand synthesis reaction was transferred to a separate tube containing 5 μCi (1000 Ci/mmol) of [α-^{32}P]-dTTP. An aliquot (usually 1 μl) of this side reaction was immediately TCA precipitated and counted to determine the background level of precipitable isotope at the start of the reaction (time 't_o'). The side reaction was incubated as for the synthesis reaction, and another 1 μl sample was TCA precipitated and counted to determine the incorporation of label at the end of the reaction (time 't_{end}'). Identical aliquots were spotted onto separate glass fibre filters and counted without TCA precipitation to determine the activity of the labelled nucleotide. The incorporation of labelled dTTP in the second strand reaction was similarly determined by adding 20 μCi (1000 Ci/mmol) of [α-^{32}P]-dTTP to the synthesis reaction and calculating the proportion of TCA precipitable material in 1 μl aliquots prior to the addition of DNA polymerase I (t_o) and at the end of the incubation (t_{end}). TCA precipitation was carried out as described in Subsection 2.4.4.1.3, but with 10% TCA.

2.4.4.3. Ligation of cDNA to UniZap XR

cDNA was ligated to the UniZAP XR vector (λZAP II vector digested with *Eco* RI and *Xho* I and phosphatased) to create the *A. cepa* cDNA expression library. For each 5 μl ligation reaction the resuspended cDNA (approximately 400 ng in 2.5 μl) was combined with 0.5 μl of 10x ligation buffer (500 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTE), 0.5 μl of 10 mM ATP, 1.0 μl (1 μg) of UniZAP XR vector and 0.5 μl (2 units) of T4 DNA ligase. Negative control reactions were identical, but lacked cDNA. Ligation reactions were incubated at 4°C for 48 hours.

2.4.4.4. Packaging

The recombinant λ bacteriophage DNA was packaged into phage heads using the Gigapack Gold packaging extracts (Stratagene). Half of each cDNA-vector ligation reaction was packaged separately. The four separate libraries obtained were pooled.

2.4.4.5. Phage assay and library amplification

The *A. cepa* cDNA expression library was titred on the *mcr*A, *mcr*B, *E. coli* host strain PLKF'. The library was titred by dilution of 2 μl aliquots with 18 μl of SM buffer. Aliquots of this diluent (usually 1-2 μl) were combined with 200 μl of prepared *E. coli* PLKF' host cells (preparation as described in Subsection 2.8.6) and incubated at 37°C for 15 minutes. The infected cells were plated as described in Subsection 2.2.3.1.

The cDNA library was amplified on the host *E. coli* PLKF' by plating at a density of approximately 20 000 pfu/plate as for the titrations. Plaques were grown at 37°C for 8 hours, prior to each plate being covered with 10 ml of SM buffer (Subsection 2.2.1.2). Plates were incubated at 4°C overnight with gentle shaking. The bacteriophage were recovered in the SM buffer from each plate and pooled in sterile 12 ml polypropylene centrifuge tubes. Chloroform was added to 5% (v/v) and the library incubated for 15 minutes at room temperature. Cell debris was pelleted by centrifugation for 5 minutes at 4 000 g. Supernatants were decanted and pooled, before being transferred to sterile glass bottles. Chloroform was added to 0.3% (v/v). The titre of the amplified library was determined as described previously in this Subsection by infection of the host *E. coli* strain XL1-Blue. The library was subsequently stored at -70°C in SM buffer with 7% dimethyl sulphoxide (DMSO).

The rate of recombination can be determined by assaying for β-galactosidase activity. To calculate the proportion of the amplified library that was recombinant, the host strain XL1-Blue was infected with dilutions of the amplified library and grown in the presence of 0.3 mM IPTG and 0.5 mg/ml X-gal. The proportion of bacteriophage capable of producing a functional β-galactosidase peptide by α-complementation as indicated by the resultant blue colour indicates the proportion of non-recombinants present within the library (Messing 1983).

2.4.5. Plaque lifts and immunoscreening of the *A. cepa* λZapII expression library

Screening of 3.2 x 10^5 pfu from the amplified *A. cepa* cDNA library was carried out on the bacterial strain *E. coli* XL1 Blue. For each plate, a dilution of the amplified cDNA library containing 2 x 10^4 pfu was added to 250 μl of XL1-Blue cells (preparation as described in Subsection 2.8.6) and incubated at 37°C for 15 minutes to allow infection. The infected bacteria were plated in 0.7% LB top agarose on 90 mm diameter sterile plastic petri plates containing LB agar. Plates were incubated at 42°C for 3.5-4 hours until plaques were just visible. Nitrocellulose filters (Hybond C, Amersham) were cut to size, soaked in 10 mM IPTG and air-dried. Dried filters were lowered on to the surface of LB plates containing the library and left at 37°C for 5 to 6 hours. Holes for later realignment of membrane and plate were made in three locations at the periphery of the membrane with a sterile syringe needle. Filters were removed from plates and washed 5 times for 5 minutes in TBST and then immunoscreened as described previously for dot blots (Subsection 2.4.2), with I66 anti-alliinase antiserum at 1:8 000 dilution and GAR-AP conjugate at 1:1 200 dilution. Wet membranes were examined over a strong light source to identify immunopositive signals and to aid in realignment of membranes with their corresponding plates.

Immunoreactive plaques and 2 non-reactive plaques (negative controls) were removed from the plates in a plug of agar and placed in 1 ml of SM buffer containing 50 μl of chloroform. The phage titre of these solutions was determined by plating serial dilutions on the host strain XL1-Blue as described (Subsection 2.4.4.5). Bacteriophage solutions were diluted and re-plated to result in 50 to 100 plaques per plate, and these membranes were re-screened with anti-alliinase antisera in an identical manner to the first round of screening. Clearly isolated positive plaques were purified by removing them in a plug of agar and transferring the agar plug to 500 μl of SM buffer with 20 μl chloroform.

2.4.6. Isolation and analysis of immunopositive clones

2.4.6.1. *In vivo* excision

The UniZAP XR vector has been designed to allow *in vivo* excision and recircularization of a phagemid containing any cloned insert within the λ vector multicloning site. The formation of the phagemid requires simultaneous infection

of an *E. coli* cell with both the λ vector and the f1 bacteriophage. DNA synthesis is initiated upstream of the multi-cloning site and terminates at a signal downstream. The single-stranded DNA molecule is circularized by a gene II product from the f1 bacteriophage, and includes all sequences of the phagemid pBluescript SK(-) and the insert DNA. This was carried out using the Stratagene protocol.

For production of double-stranded DNA, the rescued phagemid (200 μl) was combined with 200 μl of XL1-Blue plating cells (prepared as described in Subsection 2.8.6) and incubated at 37°C for 15 minutes. Aliquots of infected bacteria containing recombinant phagemid (10 μl and 100 μl) or the negative control uninfected bacteria (200 μl) were plated on LB agar supplemented with 50 μg/ml ampicillin. Ampicillin-resistant colonies were picked and re-streaked on LB agar supplemented with 50 μg/ml ampicillin.

2.4.6.2. Purification of phagemid DNA

Phagemid DNA was purified by a modified method of Colman *et al.* (1978) using hydroxyapatite column chromatography. Hydroxyapatite selectively and reversibly retains small double stranded DNA molecules under conditions known to prevent the binding of all RNA and protein. Recombinant bacteria were grown overnight in 500 ml of LB with ampicillin at 50 μg/ml and harvested by centrifugation at 6 000 g for 15 minutes. Pelleted cells were resuspended in 10 ml of 25% sucrose, 50 mM Tris-HCl pH 8.0. To prepare a cleared lysate, the resuspended cells were first incubated with 2 ml of a 10 mg/ml solution of lysozyme for 5 minutes on ice, then combined with a RNase A solution (4.0 ml of 0.25 M EDTA pH 8.0 and 20 μl of 10 mg/ml RNase A) and incubated at room temperature for 30 minutes. Lysis solution (16 ml of 0.5% (w/v) Triton X-100, 62.5 mM EDTA pH 8.0, 50 mM Tris-HCl pH 8.0) was added and gently mixed to combine. Cell debris was removed by centrifugation at 27 000 g for 75 minutes at 4°C and the supernatant was precipitated with 1/3 volume of 3 M sodium acetate pH 4.8, and 2 volumes of ethanol. DNA was collected by centrifugation at 12 000 g and dissolved thoroughly in 15-20 ml of urea/phosphate buffer (8 M urea in 0.5 M sodium phosphate buffer (0.25 M NaH_2PO_4, 0.25 M Na_2HPO_4)).

A 5-6 cm column consisting of 2 g of hydroxyapatite in urea/phosphate buffer was constructed. Sephadex G-50 (2 ml) was first pipetted on top of the sintered glass filter at the base of the column to act as a plug. Hydroxyapatite, which was prepared by washing four times with 0.5 M sodium phosphate buffer, was pipetted gently on top of the Sephadex G50 plug and allowed to pack overnight. The column was washed by gravity elution with urea/phosphate buffer for 2 hours prior to the

DNA solution being loaded onto the column by gravity feed. Urea/phosphate buffer was then washed through the column until the absorbance of the eluent, measured at a wavelength of 260 nm, was less than 0.05. Urea was removed from the column by washing with 10 mM sodium phosphate buffer for about 30 minutes. Fractions (1 ml) containing phagemid DNA were eluted from the column with 300 mM sodium phosphate buffer. The concentration and purity of the phagemid DNA in the fractions was determined spectrophotometrically (Sambrook *et al.* 1989). Fractions containing phagemid DNA were pooled and dialysed against 3 changes of $T_{10}E_1$ for 6 hours and then precipitated with ethanol as described in Subsection 2.8.5. The phagemid DNA was collected by centrifugation at 12 000 g for 15 minutes, dried under vacuum and resuspended in $T_{10}E_1$ to a concentration of between 1 and 5 mg/ml.

2.4.6.3. Restriction endonuclease digestion of purified phagemid DNA

To determine the size and relatedness of the cDNA inserts, purified phagemid DNA was digested with the restriction enzymes *Xho* I and *Eco* RI both individually and in combination. Phagemid DNA (2 μg) was diluted in the buffer supplied with the enzymes as recommended by the commercial supplier, and digested to completion in 1 hour at 37°C with 5-10 units of restriction enzyme in a 20 μl reaction volume. To restrict the phagemid with both *Eco* RI and *Xho* I, a 20 μl reaction was carried out for 1 hour with 5-10 units of each enzyme in KGB buffer (25 mM Tris acetate pH 7.5, 100 mM potassium glutamate, 10 mM magnesium acetate, 50 μg/ml BSA, 0.5 mM β-mercaptoethanol) at 37°C for 1 hour (Sambrook *et al.* 1989). A 10 μl aliquot of each digestion was separated on 1.2% agarose minigels (Subsection 2.8.1) and fragment sizes were calculated by comparison to the migration of commercial 1 kb ladder size markers (Gibco BRL).

2.4.6.4. DNA Sequencing

Double stranded phagemid templates were sequenced from both directions using the dideoxy-chain termination method (Sanger *et al.* 1977) and Sequenase as described by the kit manufacturer (USB). Sequencing from the vector/insert boundaries was carried out with the M13 Forward (5'-GTTTTCCCAGTCACGAC-3') and the SK (5'-TCTAGAACTAGTGGATC-3') primers (Stratagene). Further sequencing was carried out using 17 nucleotide-long oligonucleotides which were designed from the DNA sequence at approximately 300-400 bp intervals as it became available (Table 2.3). Synthetic oligonucleotides, were supplied in powdered form

by the manufacturer (Oligos etc.) were resuspended to a concentration of 200 μM in sterile water. Sequencing was performed as described in Subsection 2.3.4.6.

Table 2.3. Oligonucleotides designed for DNA sequencing

Oligonucleotide	Sequence
SC1	5′ GATGTTGCCAGTGGTGA 3′
SC2	5′ CCTCTTTGCTTCAACCA 3′
SC3	5′ CGACATGCTTGCCATCA 3′
SC4	5′ TATCACCTCAGCCTTGG 3′
SC5	5′ CACTTGACTGTAGAAGC 3′
SC6	5′ CCGGTGTTTAGAGAACA 3′
SC7	5′ AATGTGTTGAGGTCGCG 3′
SC8	5′ AGCTCCATGACAATACG 3′
SC9	5′ CTAAAGCGGGGAAACAA 3′

2.4.6.5. Computer-aided sequence analysis

Sequence data were aligned, searched for open-reading frames and translated to protein sequences using the IBM-PC compatible programs DM (D.W. Mount & B. Conrad, University of Arizona, Tucson USA) or Genesys (C.Bottomley, CSIRO, Canberra, Australia). Hydropathy analysis was carried out using the program SeqAid version 3.5 (DJ Routa & DD Rhoads, Division of Biology, Kansas State University, Kansas USA) and Genesys. Analysis of derived protein sequences was with the programs PepPlot, Plotstructure and IsoPept from the University of Wisconsin Genetics Computer Group (UWGCG) software package (Devereux *et al.* 1984). The PROSITE database was consulted for comparison of consensus sequences flanking known PLP binding sites (Bairoch 1992). Homology searches of the GenBank, GenPept, EMBL and SwissProt databases used the algorithms FASTA (Pearson & Lipman 1988) and BLAST (Altschul *et al.* 1990).

2.4.6.6. Expression of an immunopositive cDNA clone in *Escherichia coli*

All cDNA clones isolated were contained within the IPTG-inducible phagemid pBluescript(-) (Stratagene) as a result of their *in vivo* excision from the λZAPII vector. The clone pAlli6 was used in studies to determine the size of the protein produced from the cloned cDNA in *E. coli*. A parallel negative control consisted of XL1-Blue which contained non-recombinant pBluescript(-). LB medium

supplemented with 50 μg/ml ampicillin was inoculated with 20 μl of host cells from previous overnight cultures containing the recombinant or the non-recombinant phagemid, and grown to mid-log phase. The *lacZ* promoter was induced by the addition of IPTG to a final concentration of 1 mM. A further control consisted of cells containing the recombinant phagemid that were identically grown but were not exposed to IPTG. Cells were cultured at 37°C for 4 hours after IPTG addition. The concentration of cells, determined spectrophotometrically at 600 nm, was approximately the same in all samples. Cells were harvested by centrifugation at 5,000 g for 5 minutes from 10 ml aliquots of cultures. Extraction of total proteins from *E. coli*, and dissolution of insoluble fractions was accomplished by boiling the cell pellets for 5 minutes in 1 ml of Laemmli sample buffer (described in Subsection 2.8.2).

Aliquots (10 μl) of each sample were loaded onto two identical 12% SDS-PAGE minigels for separation of proteins as described in Subsection 2.8.2. Gels were run at 15 mA until the bromophenol blue marker dye had migrated to the bottom of the plates. One gel was used to transfer proteins to a nitrocellulose membrane by electroblotting (Subsection 2.8.4). This was subsequently probed with anti-alliinase antibodies as described in Subsection 2.4.2. The other gel was silver stained as described in Subsection 2.8.3.

2.5. Genomic library screening with an alliinase cDNA probe

Duplicate filter lifts were made from 20 nine centimetre petri dishes containing the *A. cepa*-λEMBL3 library as described in Subsection 2.2.3.1. Filters were floated on 2x SSC until wet, and submerged for 5 minutes. Filters were then incubated for two hours in prewash solution (5x SSC, 0.5% SDS, 1 mM EDTA pH 8.0) at 42°C. Bacterial debris were removed from filters by wiping gently with tissue paper wetted in prewash solution. Membranes were placed back-to-back in small plastic hybridization bags (two per bag) with 4 ml of prehybridization solution (6x SSC, 0.25% nonfat dried milk). Membranes were prehybridized for two hours at 65°C.

Hybridization was performed by incubating membranes overnight at 65°C in pre-hybridization solution with labelled probe at a specific activity $\rangle$ 10^6 cpm/ml. Membranes were washed in 2x SSC + 0.1% SDS at 65°C for 30 minutes, rinsed at the same temperature in 1x SSC + 0.1% SDS, blotted dry and wrapped in plastic

cling film. Autoradiographs were produced by exposing membranes to Kodak X-omat AR film with intensifying screens for 3 days at -70°C.

Plaques that aligned with positive signals on both replicate filters were selected and replated at a density of 50-100 pfu per plate and subsequently rescreened in an identical manner to primary screening.

2.5.1. Probe preparation and radiolabelling

Probe for this experiment was produced by cleaving the cDNA clone pAlli9 with the restriction endonucleases *Xho* I and *Eco* RI. Five μg of phagemid DNA was incubated at 37°C for 2 hours with 5 units of *Xho* I and 5 units of *Eco* RI in a 20 μl reaction volume in KGB buffer (described in Subsection 2.4.6.3). The digestion products were separated on a 1% low melting point agarose minigel in 1x TBE at 8-10 volts/cm, and visualized as described in Subsection 2.8.1. The 1.4 kb fragment cleaved from the alliinase cDNA insert was excised from the gel in a slice of agarose. This probe was labelled with [α-^{32}P]-dATP (3000 Ci/mmol) by random priming with the Amersham Megaprime kit. Unincorporated nucleotides were removed by Sephadex G50 spin-column chromatography (Sambrook *et al.* 1989). A specific activity greater than 10^8 cpm/μg was confirmed by liquid scintillation counting of 1 μl of the column eluent in a toluene-based scintillation fluid (toluene/PPO/POPOP).

2.6. Southern analysis of *A. cepa* genomic DNA

2.6.1. *A. cepa* DNA isolation and digestion with restriction endonucleases

A. cepa DNA was isolated from two bulbs of the cultivar Pukekohe Longkeeper by a modified version of the CTAB extraction method of Lassner *et al.* (1989) as described in Subsection 2.3.2.1.

Restriction enzyme digestions were carried out with 20 μg of *A. cepa* DNA in 40 μl of the appropriate 1x reaction buffer and between 10 and 20 units of restriction enzyme. Digestion reactions were incubated for nine hours or overnight at the temperature recommended by the manufacturer. Restriction enzymes used were *Bam* HI, *Bgl* II, *Cla* I, *Dra* I, *Eco* RI, *Eco* RV, *Hin*d III, *Kpn* I, *Nru* I, *Pst* I, *Sal* I, *Xba* I and *Xho* I. After incubation, a small aliquot (2 μl) of the reactions were separated

on 1% agarose gels as described in Subsection 2.8.1, to determine the extent of digestion. Digested DNA was then ethanol precipitated (Subsection 2.8.5) and dried *in vacuo*. Dried pellets were resuspended in 10 μl of $T_{10}E_1$ and digested DNA from four identical reactions pooled. Concentrations of the resuspended DNA samples were determined spectrophotometrically (Sambrook *et al.* 1989), and the volume containing either 20 μg or 60 μg was diluted to a volume of 40 μl. *P. sativum* DNA (cultivar Almota) was digested with *Hind* III and treated in the same way as *A. cepa* DNA.

2.6.2. Probes and probe labelling

Probes were created by PCR amplification of the cDNA insert within the phagemid pAlli6. A full-length cDNA probe, Alli6, was amplified using pUC/M13 universal forward (5'-GTTTTCCCAGTCACGAC-3') and reverse (5'-CAGGAAACAGCTATGAC-3') pUC/M13 17mer sequencing primers. The Alli6 probe contained the entire 1641 bp Alli6 cDNA sequence and 136 bp of flanking vector sequence. A 276 bp probe, Alli6-3', contained 202 bp of Alli6 sequence including the 3'-untranslated region (UTR), an 18 bp polyadenylate tail, and 74 bp of vector sequence. The Alli6-3' probe was amplified using the M13/pUC universal pUC/M13 forward primer and a synthetic oligonucleotide, IAL2 (5'-TTCCCGCCGTCCTTTCATT-3') designed from position 1554 to 1572 of the alliinase cDNA sequence. IAL2 was supplied as a powder by a commercial supplier (Oligos etc.) and was resuspended in water to a concentration of 200 μM.

PCR amplifications were carried out in a 100 μl reaction volume employing standard buffer conditions as described in Subsection 2.3.3. The pAlli6 template was used at a concentration of 1 ng for a 100 μl reaction. The PCR sequence consisted of a single cycle of 95°C for 6 minutes followed by 30 cycles of 95°C for 1 minute, 55°C for 1 minute and 72°C for 2 minutes, and a final soak cycle of 6°C. A Millipore spin filter with a nominal molecular mass cutoff of 30 000 Da was used to remove unincorporated dNTPs and primers from the reaction.

Probe DNA for each blot (30-50ng) was labelled with 50 μCi [α-^{32}P]-dCTP (3000 Ci/mmol) by random priming using the Amersham Megaprime kit. Alternatively an 84 μl labelling reaction was performed containing 30-50 ng of probe DNA, 0.13 M HEPES pH 6.6, 13 μM dATP, dTTP, dGTP, 33 mM Tris-HCl pH 8.0, 3 mM $MgCl_2$, 6.5 mM β-mercaptoethanol, 0.4 mg/ml BSA, 3 units DNA polymerase (Klenow fragment), 50 μCi [α-^{32}P]-dCTP (3000 Ci/mmol), 0.27 O.D. units (1.1 μg) of random hexamers (Pharmacia) (Feinberg & Vogelstein 1983). The reaction was

incubated for 2 hours and halted by the addition of 16 μl of STE (0.1 M NaCl, 10 mM Tris HCl pH 8.0, 1 mM EDTA pH 8.0). Unincorporated label was removed from probes by Sephadex G50 spin-column chromatography (Sambrook *et al.* 1989). A specific activity of greater than 10^8 cpm/μg of probe DNA was confirmed after each reaction by scintillation counting of a 1 μl aliquot of the column eluent in a toluene-based scintillation fluid (toluene\PPO\POPOP).

2.6.3. Southern blotting and hybridization conditions

DNA samples for Southern analysis (Southern 1975) were separated on 1.2% agarose gels in 1x TBE overnight at 3 volts/cm. Standard 1 kb DNA molecular size markers (Gibco BRL) were loaded on all gels. Gels were stained briefly and DNA was examined as described in Subsection 2.8.1. Southern transfer and hybridization was carried out as described in Subsection 2.3.2.3. Membranes were washed in 2x SSC, 0.1% SDS for 30 minutes, or sequentially in 2x SSC, 0.1% SDS, followed by 1x SSC, 0.1% SDS and 0.1x SSC, 0.1% SDS for 30 minutes each. For re-probing, membranes were stripped of bound probe by washing in 0.1 N NaOH for 15 minutes at room temperature. Membranes were then neutralized in 0.1 M Tris-HCl pH 7.5, and washed in 0.1 M Tris-HCl pH 7.5, 2x SSC, 2% SDS for 10 minutes at 42°C, followed by a wash in 0.1x SSC, 0.1% SDS for 10 minutes at 42°C (Sambrook *et al.* 1989). Prior to reuse, stripped membranes were examined for any residual probe binding by exposure to Kodak X-omat AR film for 4 days with intensifying screens at -70°C.

Southern blot membranes probed with the Alli6 cDNA probe were exposed to Kodak X-omat AR film for 1 week with intensifying screens at -70°C. Those probed with the Alli6-3' probe were exposed for 2 weeks.

2.6.4. Copy number reconstructions

The 1C genome size of *Allium cepa* has been estimated to be 1.55 x 10^7 kbp (Arumuganathan & Earle 1991). A single copy of a 1.7 kbp gene therefore represents 1.097 x 10^{-5}% of the genomic DNA present, equal to 2.19 pg in 20 μg of DNA. Single copy reconstructions were made from dilutions of Alli6 probe DNA and were loaded on all gels used for Southern analysis.

2.7. Northern analysis of *A. cepa* seedling RNA

2.7.1. Sample selection

Seeds of the *A. cepa* cultivar Pukekohe Longkeeper were sterilized in a 1:20 dilution of 3% (w/v) sodium hypochlorite for 30 minutes followed by 5 rinses of sterile distilled water then placed on two layers of wet Whatman 3MM paper in small covered trays. Half of the trays were wrapped in two layers of aluminium foil to exclude light. Seeds were germinated and seedlings grown at 22°C.

Seedlings were selected from five growth stages which were based on a previously described scale (McCallion & Lancaster 1984). Stage 1, root emerged (about 1-2 days after imbibition; 1-2 DAI); stage 2, hypocotyl emerged (about 2-3 DAI); Stage 3, early developing cotyledon (about 3-6 DAI); stage 4, elongating cotyledon, adventitious roots emerging (about 6-9 DAI); Stage 5, emerging first leaf (about 10-14 DAI). Collection was halted at 14 DAI as dark-grown seedlings did not survive past this time. Seed coats were manually removed from *A. cepa* tissue after it was frozen in liquid nitrogen. Collection of dark-grown tissue was carried out in a darkroom under a green Kodak safe light.

2.7.2. RNA extraction

RNA was extracted from 1-2 g of tissue by the method of Logemann *et al.* (1987) as described in Subsection 2.4.4.1.1. RNA was also extracted from *Solanum tuberosum* (potato) meristem tissue for use as a negative control. RNA was quantitated and its purity determined spectrophotometrically as described in Subsection 2.4.4.1.1 (Sambrook *et al.* 1989), and 10 μg of each sample was fractionated on 2% formaldehyde/1% agarose gels in 1x MOPS buffer at 3 volts/cm overnight as described in Subsection 2.4.4.1.1. Molecular size markers were 1 μg of a 1 kb RNA ladder (Gibco BRL). Formaldehyde was removed from the gel by soaking in two changes of 10x SSC for 20 minutes each. RNA was transferred to Zetaprobe nylon membranes (Biorad) by capillary transfer in 20x SSC (Thomas 1980). Membranes were then rinsed in 6x SSC for 5 minutes and air-dried.

The integrity of the RNA samples transferred to membranes was confirmed by subsequent visual examination of agarose gels identical to those used for northern blotting. RNA samples were visualized as described in Subsection 2.4.4.1.1.

2.7.3. Hybridization conditions

Membranes were washed in 2x SSC for 15 minutes at 65°C, then prehybridized as described in Subsection 2.3.2.3 for 4 hours at 65°C in heat-sealable plastic bags. The labeled probe mixture was then added and hybridizations were carried out overnight at 65°C with agitation. Hybridization solutions were removed and membranes were washed at 65°C as described in Subsection 2.6.3. Membranes were then blotted dry and sealed in plastic wrap. Northern blot membranes were exposed to Kodak X-omat AR film for 4 to 5 hours at -70°C with intensifying screens.

2.7.4. Probes and probe labelling

Probes for northern blot membranes were prepared as described in Subsection 2.6.2. A *Malus* (crab apple genotype 'White Angel') 45S rDNA clone containing coding regions for 5.8S, 18S and 26S rRNAs (Simon & Weeden 1992) was kindly provided by Dr S. Gardiner (Hort+Research Ltd., Palmerston North). To estimate variations in RNA loading and transfer to membranes, all northern blot membranes were probed with this 45S rDNA.

2.7.5. Densitometer scanning

Autoradiographs from blots were scanned using a Shimadzu Model CS-930 densitometer with a 1 cm wide window at a wavelength of 540 nm. Scans were repeated, and the integrated optical density (IOD) values obtained for each lane were averaged. IOD values of signals produced by northern blots probed with the Alli6 cDNA probe were normalized relative to IOD values obtained for the entire lane from hybridization of the same membranes with a *Malus* 45S rDNA probe. IODs obtained from probing northern blots with ribosomal DNA probes are linearly related to the amount of RNA loaded for amounts between 2-30 μg (Correa-Rotter *et al.* 1992). Four replicate northern blots were analysed to correct for loading and transfer variation. IOD data obtained from the replicate northern blots was analysed by analysis of variance (ANOVA).

2.8. Miscellaneous methods

2.8.1 Separation of DNA fragments by agarose gel electrophoresis

The sizes of various DNA species were estimated by gel electrophoresis. Agarose gels were made in 1x TBE. Samples were combined with 0.2 volumes of sample buffer (50% glycerol, 0.25% bromophenol blue in 5x TBE) and loaded into each lane. Molecular size markers were 1 μg of 1 kb ladder or *Hin*d III-digested λ DNA (Gibco BRL) and electrophoresis was in 1x TBE at between 3 and 5 volts/cm. After the blue dye had migrated to between one half and two thirds of the gel length, gels were stained in a 0.5 μg/ml solution of ethidium bromide for 15 minutes. Gels containing Nusieve were destained in water for 20-40 minutes. Gels were examined on a short wavelength ultraviolet transilluminator.

2.8.2. SDS polyacrylamide gel electrophoresis

Protein samples were mixed with an equal volume of Laemmli sample buffer (62.5 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 5% (v/v) β-mercaptoethanol) and denatured by boiling for 3 minutes. Bromophenol blue was added to 0.0005% (w/v) just prior to loading samples onto gels. Electrophoresis was performed on a 12% polyacrylamide separating gel (375 mM Tris-HCl pH 8.8, 12% acrylamide:bis (29.2:0.8), 0.1% SDS, 0.05% ammonium persulphate, 0.33% (v/v) TEMED) with a 4% polyacrylamide stacking gel (0.12 M Tris-HCl pH 6.8, 4% polyacrylamide:bis (29.2:0.8), 0.1% SDS, 0.05% ammonium persulphate, 0.05% TEMED) in electrophoresis buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% SDS) according to the method of Laemmli (1970). For gels that were subsequently silver stained, the Biorad low-range molecular mass protein markers used were phosphorylase b (97 400 Da), bovine serum albumin (66 200 Da), ovalbumin (42 699 Da), bovine carbonic anhydrase (30 000 Da), soybean trypsin inhibitor (21 500 Da) and lysozyme (14 400 Da). For gels which were to be used for electroblotting, the Amersham Rainbow molecular mass markers used were myosin (blue, 200 000 Da), phosphorylase b (brown, 92 500 Da), bovine serum albumin (red, 69 000 Da), ovalbumin (yellow, 46 000 Da), carbonic anhydrase (orange, 30 000 Da), trypsin inhibitor (green, 21 500 Da), lysozyme (magenta, 14 300 Da). Gels were run at 15 mA until marker dye migrated to the bottom of the plates.

2.8.3. Silver staining

Gels were fixed for 30 minutes in 50% (v/v) ethanol, 10% (v/v) acetic acid, then transferred to 10% glutaraldehyde for 30 minutes. After extensive washing in distilled water for at least 2 hours, gels were stained in freshly made ammonical silver-stain solution (18.9 mM NaOH, 207 mM NH_4OH, 45.6 mM $AgNO_3$) for 3 minutes with agitation. Gels were then washed twice with distilled water and immersed in freshly prepared developing solution (0.24 mM citric acid, 0.0185% formaldehyde) for 5 minutes. Gels were washed again with water and placed in either 50% methanol, or 45% methanol/10% acetic acid to halt the stain development.

2.8.4. Electroblotting

Proteins were transferred from SDS-PAGE minigels to nitrocellulose membranes using a LKB Novablot semi-dry electroblotting apparatus. The SDS-PAGE gel and a nitrocellulose filter of identical size were sandwiched between eight pieces of Whatmann 3MM paper moistened in transfer buffer (39 mM glycine, 48 mM Tris base, 1.3 mM SDS, 20% (v/v) methanol). Electro-transfer was at 0.8 mA/cm^2 for 1 hour.

2.8.5. Ethanol precipitation

Various DNA samples were concentrated by precipitation with 70% ethanol in the presence of 0.3 M sodium acetate pH 5.2. Precipitations were carried out either overnight at -20°C or for 15 minutes at -70°C. DNA was collected by centrifugation at 12 000 g for 10 minutes at 4°C or at room temperature.

2.8.6. Preparation of plating cells

Bacterial host strains were prepared for infection by λ bacteriophage by inoculating 10 ml of LB broth supplemented with 0.2 % maltose and 10 mM $MgSO_4$, with a fresh colony grown on LB plates which were supplemented with 12.5 μg/ml tetracycline for the *E. coli* strains XL1-Blue and PLKF'. Liquid cultures were shaken overnight at 30°C. Cells were collected by centrifugation at 1 000 g for 10 minutes, then gently resuspended in 10 mM $MgSO_4$ to an optical density of 0.5 at a wavelength of 600 nm.

Chapter 3. Results

3.1. Protein sequence analysis of alliinase from *A. cepa*

The primary objective of this study was to isolate an alliinase clone. To this end alliinase protein sequence analysis was undertaken. It was anticipated that the sequence obtained would then be used to design oligonucleotide probes which could select clones from gene libraries and/or confirm clone identity by hybridization analysis. Alternatively oligonucleotides could be used to prime PCR amplifications from *A. cepa* cDNA or genomic DNA templates as a means to isolate a partial alliinase clone. Protein sequence data could also ultimately serve to confirm clone identity by alignment with a translation of putative alliinase clone open reading frames (ORFs). *A. cepa* alliinase protein was purified by M.L. Shaw (Crop and Food Research Ltd., Lincoln). Sequences were initially obtained from the N-terminus of intact alliinase and later from internal peptide fragments generated by cleavage with cyanogen bromide (see Subsection 2.1). These sequences are presented in Table 3.1. Two sequences were ascertained from intact native alliinase, the 23 residue sequence ALN1, and the 17 residue sequence, ALN2. ALN1 differed in only one residue from the four N-terminal amino acids, KMTW, of the *A. sativum* alliinase determined by Nock & Mazelis (1989). Searches of the SwissProt protein sequence database later revealed that ALN2 showed high identity to the lectins of *A. sativum* (van Damme *et al.* 1992) and was consequently thought to have most likely arisen by co-purification of *A. cepa* lectin with alliinase. No homologies were found to the other peptide sequences obtained.

Table 3.1. N-terminal amino acid sequences from native and cyanogen bromide-cleaved fragments of purified *A. cepa* alliinase[1]

N-terminus of mature alliinase
ALN1. KVTWSLKAAE(C)AEAVANINQSGF
ALN2. N(?)LVNNEGLY(A)G(?)(?)LVE
N-terminus of cyanogen bromide-cleaved fragments of alliinase
ALN3. AYFFNPVSNFISFE
ALN4. VYYWPHYTPIKYKADED
ALN5. YYLKN

[1] Alliinase was purified from the *A. cepa* cultivar Pukekohe Longkeeper.
( ) = uncertainly determined residues.

3.2. Genomic library screening with redundant oligonucleotides

3.2.1. Genomic library construction

At the time this project was initiated, amino acid sequence analysis of native *A. cepa* alliinase was in progress, and an *A. cepa* root tip library in λgt10 had been constructed by J.E. Lancaster and J. Farrant (Crop and Food Research Ltd., Lincoln). It was anticipated that alliinase clones within this cDNA library would be identified by hybridization analysis using redundant oligonucleotides designed by back-translation of the forthcoming alliinase peptide sequence. These could then be used as alliinase gene-specific probes. To complement the work already underway, this project initially focused on isolating a genomic clone of alliinase. A genomic clone bank of *A. cepa* was created by ligating DNA fragments with a predominant size range between 16 and 25 kb to the λ bacteriophage vector, EMBL3 (see Subsection 2.2.1).

A. cepa DNA for genomic library construction was extracted by the method of Kirk *et al.* (1970) whereby nuclei are first isolated and then lysed to release nucleic acids. Restriction enzyme digestion conditions were optimized across a range of DNA and *Sau* 3A concentrations (see Subsection 2.2.1.1.2). Fractionation of 500 μg of digested *A. cepa* DNA on NaCl gradients by ultracentrifugation yielded approximately 5 μg of DNA in the appropriate size range for library construction.

λEMBL3 vector arms to *A. cepa* DNA insert molar ratios were varied around the theoretical ideal of 2:1 to ascertain the optimum conditions for ligation (see Subsection 2.2.1.2). Altering the vector arms to insert concentrations across the molar ratios between 9:1 to 0.9:1 did not make a significant difference in the resulting number of pfu per μg of insert DNA (Table 3.2). Efficiency was low as compared to reactions containing commercially supplied insert DNA and was not sufficient to produce a library representative of the entire *A. cepa* genome. Consequently a clone bank was constructed which consisted of the several small gene banks created while optimizing library construction conditions, plus several further ligation and packaging reactions. In total this library resulted from ligating 3.36 μg of fractionated *A. cepa* DNA to λEMBL3 arms and contained 2.44 x 10^5 pfu, equivalent to 5% of the size theoretically required to represent the entire *A. cepa* genome.

Table 3.2. Effect of altering the ratio of EMBL3 vector arms to *A. cepa* insert DNA on the efficiency of library construction[1]

Insert (ng)	Vector Arms (ng)	Vector:Insert Molar Ratio	Plaques formed per μg of insert DNA
20	500	8.75:1	5.0×10^4
40	500	4.4:1	4.1×10^4
120	500	1.5:1	2.0×10^4
200	500	0.9:1	5.2×10^4
pTI11 test 87.5	500	2.0:1	1.2×10^8
none	500	-	4.0×10^2

[1] Average values from two duplicate trials.
Vector DNA, 500 ng = 0.0175 pmole. Insert DNA of 15 kb, 100 ng = 0.01 pmole.

3.2.2. Genomic library screening with oligonucleotides

It was originally intended that genomic library screening would be carried out by hybridization with cDNA probes. However difficulties were encountered in isolating alliinase clones from the *A. cepa* λgt10 library (J. Farrant pers. comm.). Consequently preliminary studies were performed to ascertain the feasibility of screening the *A. cepa* genomic library with redundant oligonucleotides designed by back-translation of the ALN1 peptide sequence. Other alliinase peptide sequences shown in Table 3.1 were not used to design probes for this study as they were not available at the time this work was performed.

The unamplified genomic library was screened by hybridization with both the JL1 and JL2 probes (Table 2.1, page 54) at 55°C. Two positive signals were detected that aligned both on duplicate filters and with plaques on their corresponding plates. These putative positive clones were selected from the plates and re-screened at 55°C in duplicate at low density (approximately 100-200 pfu per plate). However, neither of these selected clones gave consistently positive signals aligning with plaques on rescreening. Therefore it appeared that isolation of alliinase clones using this approach might not be feasible.

3.3. Preliminary studies on PCR amplification of alliinase-encoding sequences with redundant oligonucleotides

PCR was investigated as a means to generate a partial alliinase-encoding fragment which could subsequently be used as a probe to isolate an alliinase clone from either cDNA or genomic gene libraries. Four approaches were taken, these were (i) amplifications from cDNA templates, (ii) 'nested' amplifications, (iii) amplifications from genomic DNA and (iv) amplifications from genomic DNA templates selected by Southern hybridization. The various reactions performed are summarized in Table 2.2, page 64.

As the *A. cepa* alliinase subunit size is 50 kDa (Tobkin & Mazelis 1979), of which between 4.6% and 5.8% is carbohydrate (Nock & Mazelis 1986, 1987) the corresponding alliinase transcript would be expected to be at least 1.2 kb. It was anticipated that using sense primers designed from the ALN1 peptide sequence in combination with $(dT)_{17}$ would amplify products from cDNA templates of greater than this size. Independent amplifications using adjacent sense primers, each in combination with a particular antisense primer, allowed predictions to be made of the expected difference in product sizes which could also be used to aid identification of putative authentic fragments. It was further anticipated that manipulation of reaction conditions might result in the amplification of a single product which could then be identified by DNA sequencing.

3.3.1. Amplifications from cDNA templates

3.3.1.1. Mixed oligonucleotide-primed amplification of cDNA

Initial studies were based on the methods of mixed oligonucleotide-primed amplification of cDNA (MOPAC) (Lee & Caskey 1989) and the rapid amplification of cDNA ends (RACE) PCR (Knoth *et al.* 1988). Amplifications were primed with the N-terminal sense primers JL5, JL8 and JL9 in combination with either $(dT)_{17}$ or the antisense primer JL4 (see Figure 2.1 and Table 2.1, pages 53 and 54). A range of primer concentrations, $MgCl_2$ concentrations and annealing temperatures were used to optimize the reaction conditions (summarized in Table 2.2). These reactions consistently produced high levels of background products which were seen as a bright smear on agarose gels. Furthermore the primer JL4 produced several spurious fragments in control reactions with no corresponding sense primer provided (data not presented).

Reactions containing the JL5 primer in combination with $(dT)_{17}$ occasionally produced a faint product of about 1.5 kb which was not observed in single-primer

control reactions. The fragment was gel-purified and used as a template for further amplifications. This template isolation in conjunction with optimization of the annealing temperature, cycle number, magnesium ion concentration and primer concentration (subsection 2.3.4.1) significantly lowered the levels of background products and resulted in the consistent amplification of a single product (Figure 3.1, page 93). Optimal conditions were found to be 30 PCR cycles, with a 60°C annealing temperature, 1 μM of each primer and 1.5 mM $MgCl_2$.

The reamplified DNA fragment was then again isolated from agarose gel slices, blunt end ligated to the plasmid vector pUC19 and transformed into *E. coli* TG1 (see Subsection 2.3.4). Colonies were initially screened for recombinant plasmids by blue/white selection on X-gal and IPTG. Further identification of transformants was by PCR with the JL5 and $(dT)_{17}$ primers using crude colony preparations as templates (described in Subsection 2.3.4.4). Of 47 white colonies screened, one contained a plasmid that yielded a 1.5 kb insert on reamplification with JL5 and $(dT)_{17}$.

This cloned PCR product was partially DNA sequenced from both insert-vector borders, and these data are presented in Figure 3.2, page 95. The cloned fragment contained one of the redundant JL5 primer sequence options at each boundary, with a single base mismatch at one end and two at the other. Translation of the DNA sequences adjoining the primer regions did not produce a peptide sequence homologous to that of alliinase, and consequently it was clear that this product was unlikely to be derived from an alliinase transcript. The arrangement of JL5 sequences at both termini suggested this product should be amplified by PCR with the JL5 primer alone, however this did not occur and $(dT)_{17}$ was found to be essential for its production (data not presented). This was thought to imply that this 1.5 kb fragment may have arisen by concatemerization of two smaller products generated during the PCR.

3.3.1.2. Nested PCR

In nested PCR an amplification is performed with a single primer pair, and an aliquot of this reaction is subsequently reamplified with a set of inner, 'nested' primers. This 'nesting' can significantly improve the specificity of the PCR, since the final product is the result of amplification with two pairs of primers (Ohara *et al.* 1989, Dietz *et al.* 1990). Reactions initially performed with the sense primers JL5 and JL8 were reamplified with the adjacent downstream sense primers JL8 and JL9 respectively (see Figure 2.1). However for each set of reactions, the same antisense primers (either JL4 or $(dT)_{17}$) were used (see Table 2.2).

Annealing temperatures were kept low (ie. between 35-45°C) during nested PCR amplifications (see Table 2.2) as it was anticipated that specificity would be

conferred by the subsequent amplification round with a different sense primer. An alliinase cDNA product amplified with the N-terminal sense primers (JL5, JL8 and JL9) in combination with $(dT)_{17}$ would be expected to be greater than 1.2 kb long. The size of authentic products resulting from amplification with the internal antisense primer JL4 could not be predicted, but should be 30 bp longer when produced in combination with the sense primer JL8 then with JL9 (see Figure 2.1).

High levels of background products were evident after these nested reactions (data not presented). When amplifications were primed with JL5, JL8 and JL9 in combination with $(dT)_{17}$, multiple bands less than 1 kb in size were observed by agarose gel electrophoresis. The JL4 antisense primer amplified numerous products either when present in reactions containing sense primers, or when alone. As the sizes of products generated by these PCRs were substantially less than predicted for alliinase cDNAs, this suggested that alliinase cDNAs may be present at low abundance, possibly due to low efficiency of the reverse transcription reaction during the cDNA synthesis step, and/or to a low abundance of alliinase transcripts within the RNA samples used. Consequently subsequent PCR experiments focused on the use of *A. cepa* genomic DNA templates.

3.3.2. Amplification from *A. cepa* DNA templates

Considerable quantities of spurious products may be amplified from genomic DNA templates when PCRs are primed with highly degenerate oligonucleotides. As no data on *A. cepa* codon usage was available at the time these studies were carried out, reliable codon choices to restrict primer degeneracy could not be made, and an alternative strategy was necessary. The oligonucleotides JL11, JL12, JL13 and JL14 were designed from four to five residues of peptide sequence with the intention of limiting primer degeneracy by reducing primer length. JL11, JL12 and JL13 were designed from the N-terminal peptide sequence, ALN1 (see Figure 2.1). Because the JL4 primer designed from residues one to nine of ALN3 had resulted in high levels of nonspecific product amplification in the previous studies (see Subsection 3.3.1.2), the primer JL14 was designed from residues ten to 13 of this internal peptide sequence. This resulted in a series of primers with between 32-fold and 192-fold redundancy (Table 2.1).

The sense primer JL11 was used in conjunction with either of the antisense primers JL12 or JL14, whereas the sense primer JL13 was used only in combination with the antisense primer JL14 (Table 2.2). As JL11 and JL12 were located at opposite ends of the ALN1 peptide sequence, this primer combination could be predicted to amplify a 60 bp fragment from cDNA templates. However the size of the product amplified from a genomic template could not be predicted. An authentic product resulting from JL11/JL14 reactions would be expected to be 33 bp larger

than a genuine product amplified with the JL13/JL14 primer pair (Figure 2.1).

These reactions resulted in the amplification of multiple fragments of various sizes less than 1.2 kb under all the reaction conditions tried (see Table 2.2). These primers also generated numerous fragments in lone-primer control reactions (data not presented). Amplification from *A. cepa* DNA templates were also performed using the primers JL5, JL8, JL9 and JL4 (see Table 2.2), and generated a large range of products in standard and nested PCRs which could not be individually resolved or isolated for further examination.

3.3.3. PCR with templates selected by Southern hybridization

To improve the specificity of amplifications, templates were enriched for alliinase-specific sequences by Southern analysis (described in Subsection 2.3.2.3) following the approach of Beck & Ho (1988). The oligonucleotides JL4, JL5 and JL8 were observed to hybridize to the digested *A. cepa* DNA in each lane as a smear, or hybridize with equivalent affinity to *A. cepa* DNA and negative control *T. aestivum* DNA fragments. However the JL9 probe showed no complementarity to *T. aestivum* DNA, and hybridized to five specific *A. cepa* DNA bands. These were *Hin*d III-generated fragments of 1.85 kb and 2.2 kb, *Eco* RI-generated fragments of 3.2 kb and 3.8 kb, and an *Eco* RV-generated fragment of 4.9 kb. Consequently *A. cepa* DNA fragments of these sizes were eluted from agarose gel slices after separation by electrophoresis, and used as PCR templates.

PCR amplifications from these templates were with the oligonucleotides JL11, JL12, JL13 and JL14 as in the previous study. The JL11/JL12 primer combination amplified a single 470 bp band from all five templates. This fragment had been observed among the multiple products amplified from unfractionated *A. cepa* genomic DNA in previous reactions (described in Subsection 3.3.2). Reactions containing both JL11 and JL14 produced a fragment of approximately 350 bp from the *Eco* RV/4.9 kb template and the *Hin*d III/1.85 kb templates, but not from the others. JL13 and JL14 in combination did not result in amplification of any products from these templates. Identical reactions in which the templates were replaced by similarly prepared *T. aestivum* DNA fragments did not yield any amplification products (data not presented).

From the alliinase peptide sequence it was predicted that the combination of the JL11 and JL12 primers should amplify a smaller fragment than the JL11 and JL14 primer combination (see Figure 2.1). However as this is opposite to the result obtained, the authenticity of these products was uncertain. In addition, polyclonal antibody screening of an *A. cepa* cDNA expression library had by this time resulted in the identification of several immunopositive clones. Therefore no further attempts were made to amplify alliinase-encoding sequences by PCR.

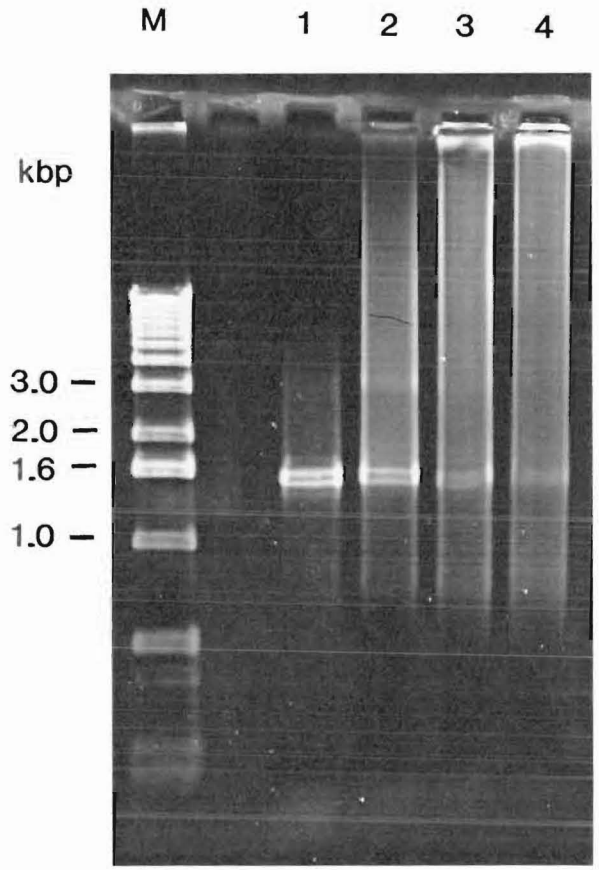
Figure 3.1. Optimization of buffer conditions and annealing temperature for reamplifying a PCR product amplified from *A. cepa* cDNA

A PCR product was purified from an agarose gel slice (see Subsection 2.3.4.2). Reactions were carried out to optimize amplification from this fragment template. Standard reactions contained 1 ng template DNA, 0.2 mM each dNTP in 1x PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin (w/v)) and 0.4 units *Taq* DNA polymerase. Standard PCR conditions consisted of 1 cycle at 95°C for 6 minutes followed by 30 cycles of 95°C for 1 minute, 37°C for 1 minute and 72°C for 2 minutes followed by a 6°C soak step.

Panel A. $MgCl_2$ concentration was varied. Lane 1; 1.5 mM, Lane 2; 2.0 mM, Lane 3; 3.0 mM, Lane 4; 4.0 mM. M; 1 kb DNA size markers.

Panel B. Annealing temperature was varied. Lane 1; 40°C, Lane 2; 50°C, Lane 3; 60°C, Lane 4; 65°C. M; 1 kb DNA size markers.

A



B

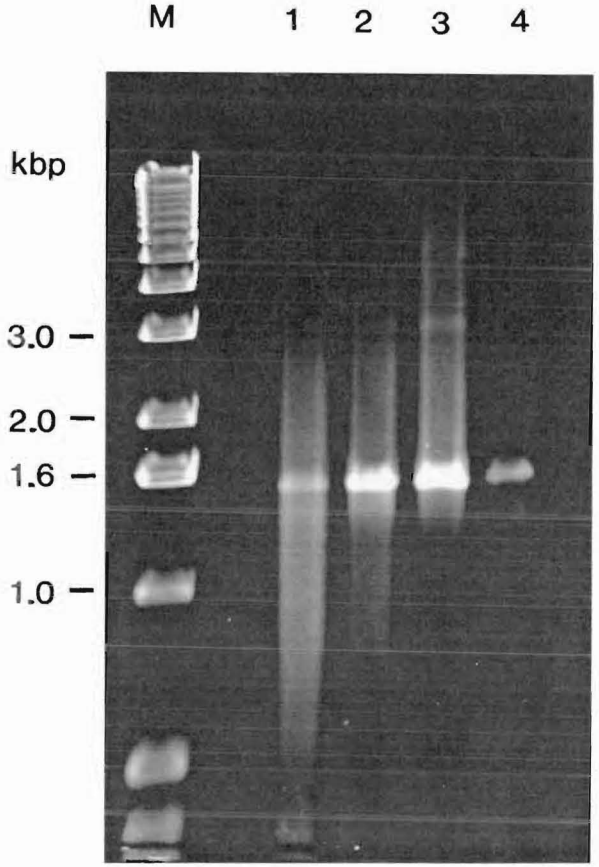


Figure 3.2. Partial sequence of a PCR-produced fragment and its predicted amino acid sequence

The partial sequence of a 1.5 kb product that was amplified by PCR with the oligonucleotides JL5 and $(dT)_{17}$ is presented. This PCR product was isolated from an agarose gel slice, cloned into pUC19 and partially sequenced from the vector/insert borders using universal pUC/M13 forward and reverse sequencing primers (see Subsection 2.3.4). The amino acid sequence shown was predicted by assuming the reading frame was contiguous with that of the theoretical alliinase mRNA. ├SalI┤ = *Sal* I restriction site. _ = unclear nucleotide. *X = nucleotides not present at that site in the JL5 primer.

A. Forward primed sequence

```
                                        Lys Val Thr Trp Thr Leu Lys Ala
5'GATTACGAATTCGAGCTCGGTACCC GGTCGACC AAA GTT ACT TGG ACG CTG AAG GCA
  ├──────────pUC19──────────┤─SalI─├──────────JL5──────────┤

 Ala Asp Lys Arg His Gln His His Leu Ala Leu Arg Ile Cys Asp Arg Ser
 GCA GAT AAA CGA CAC CAG CAT CAC CTG GCG CTT CGC ATT TGT GAC CGC AGT

 Lys Ser Gly Glu Gln Val Ile Ala Ala Val Asn Gln Ile Ala Asp Asp His
 AAA AGC GGC GAG CAG GTC ATA GCA GCC GTA AAT CAG ATA GCT GAC GAC CAC

 His Pro Asp Arg Thr     Ser Ala Thr Ile Val Val
 CAC CCC GAC CGC ACT AA_ AGC GCA ACG ATT GTA GTT 3'
```

B. Reverse primed sequence

```
                                        Lys Val Thr Trp Ser Phe Lys Ala
5'TGCAGGTCGACTCTAGAGGATCCCC GGTCGACC AAG GTG ACG TGG TGG TTT AAG GCG
  ├─────────pUC19──────────┤─SalI─├──────────JL5──────────┤

 Ile Gly Glu Ala Gln Glu Arg Ile Ile Leu Glu Thr Phe Ile Trp Phe Glu
 ATT GGC GAG GCA CAA GAA CGC ATC ATT CTT GAA ACG TTT ATC TGG TTT GAG

Asp Asp Val Gly Lys Gln Leu His
GAT GAC GTC GGC AAA CAA CTG CAT 3'
```

3.4. Immunoscreening of an *A. cepa* cDNA expression library

3.4.1. Construction of an *Allium cepa* sprouting leaf cDNA library

An expression library was constructed in λZAPII with cDNA synthesized from *A. cepa* (cultivar Southport White Globe) poly(A)+ RNA. RNA for cDNA library construction was extracted from sprouting *A. cepa* bulb inner leaves as this tissue had been found to yield abundant quantities of RNA (see Subsection 2.4.4.1.1). Examination of the RNA samples by agarose gel electrophoresis showed well-defined 28S and 18S ribosomal RNA bands and a smear of other RNA species indicating the RNA was largely undegraded. Fractionation of 5 mg of total *A. cepa* RNA using the PolyATtract extraction system (Promega) yielded between 4.8 μg and 5 μg of poly(A)+ RNA.

To confirm the integrity of the poly(A)+ RNA isolated, *in vitro* translation experiments were carried out using a rabbit reticulocyte lysate system (see Subsection 2.4.4.1.3). *In vitro* translation of *A. cepa* poly(A)+ RNA resulted in a 2.2-fold increase in TCA-precipitable counts over background levels (Table 3.3). Examination of the translation products after separation by SDS-PAGE revealed a large number of proteins of between 20 kDa and 90 kDa (Figure 3.3, page 101), indicating that the *A. cepa* poly(A)+ RNA was substantially intact and included transcripts of more than 2 kb. On the basis of this evidence, the poly(A)+ RNA was believed to be of a quality suitable for cDNA library construction.

Table 3.3. Incorporation of ^{35}S-methionine into protein by *in vitro* translation of *A. cepa* poly(A)+ RNA

Sample	Time (minutes)	TCA precipitable cpm per reaction	-fold increase
No RNA (Negative control)	0	5.7×10^5	1.00
	60	5.7×10^5	
***A. cepa* poly(A)+ RNA**	0	5.8×10^5	2.24
	60	1.3×10^6	
Rabbit globin mRNA (Positive control)	0	6.5×10^5	3.7
	60	2.4×10^6	

The cDNA was synthesized, and the cDNA library was created as described by the manufacturers of the Stratagene ZAP-cDNA synthesis kit (see Subsection 2.4.4.2). The efficiencies of cDNA synthesis reactions were calculated by the incorporation of [α-^{32}P]-dTTP into first and second strand cDNA as determined by TCA precipitation and scintillation counting. First strand synthesis resulted in the production of 1.2 μg of cDNA, equivalent to a 26% conversion of the RNA supplied in the reaction. Second strand synthesis resulted in the production of 830 ng of double-stranded cDNA, equivalent to a conversion of 70% of the first strand cDNA template to double-stranded cDNA. These data are presented in Table 3.4.

The *A. cepa* cDNA (800 ng) with *Eco* RI and *Xho* I linkers was ligated to 2 μg of UniZAP XR vector arms and packaged (described in Subsections 2.4.4.3 and 2.4.4.4). The resultant library contained a total of 2.9 x 10^5 pfu as titred on the Mcr- *E. coli* host strain PLKF'. Therefore the efficiency of library construction (ie. ligation and packaging) was calculated at 3.5 x 10^5 pfu/μg of cDNA. The library was amplified to a titre of 2.2 x 10^{10} pfu/ml on the *E. coli* host strain XL1-Blue (Subsection 2.4.4.5). Approximately 53% of plaques from the amplified library were unable to metabolize Xgal (Lac-), evidence that more than half of the bacteriophage contained within the library were recombinants. Packaged negative control reactions (ie. ligations lacking insert DNA) produced an acceptably low background of blue plaques.

Table 3.4. Yield and reaction efficiencies for first and second strand cDNA synthesis as measured by incorporation of [α-^{32}P]-dTTP

	Proportion of isotope incorporated (%)[1]	Yield (μg)	Efficiency (%)
First strand	4.06	1.20	26[a]
Second strand	0.60	0.83	70[b]

[1] proportion of counts in TCA precipitable fraction at t_{end} minus the proportion of counts in TCA precipitable fraction at t_0.
[a] yield as a percentage of poly(A)+ RNA template.
[b] yield as a percentage of first strand cDNA template.

3.4.2. Production of anti-alliinase polyclonal antibodies

Anti-alliinase antibodies were induced in two rabbits, I66 and I65 against alliinase purified as described in Subsection 2.1. The antiserum was titred for the activity of anti-alliinase antibodies using ELISA techniques (described in Subsection 2.4.1). The I66 antiserum was found to be reactive to 200 ng of *A. cepa* alliinase when diluted 1:80 000, while the I65 polyclonal sera could not detect 200 ng alliinase at dilutions beyond 1:16 000. As signal detection is improved with polyclonal serum of higher titre (Helfman & Hughes 1987), the I66 polyclonal serum was employed in all immunological detection studies.

Polyclonal preparations can contain conformation-dependent antibodies which may bind non-specifically to surfaces such as the membrane being used, or may contain IgG components that bind to *E. coli* proteins. It is therefore important to determine the greatest dilution of antibody that results in a strong positive reaction and whether the primary antibody reacts with bacterial proteins which would increase the background and produce false positives. The optimal levels of I66 serum and GAR-AP conjugate for library screening were determined by probing dot blots containing dilutions of purified *A. cepa* alliinase, dilutions of anti-alliinase serum, and dilutions of *E. coli*/bacteriophage lysate (see Subsection 2.4.2). An anti-alliinase antiserum dilution of 1:8 000 and a secondary antibody concentration of 1:1 200 were found to produce the optimal signal to background ratio and were consequently used for all immunoscreening and western analysis.

To determine whether the alliinase protein back bone contained epitopes recognized by the anti-alliinase antibodies, purified *A. cepa* alliinase was stripped of its oligosaccharides with TFMS then western blotted and probed with I66 anti-serum (see Subsection 2.4.3). The I66 anti-alliinase antibodies exhibited a significant affinity for the deglycosylated alliinase although the colour reaction against the intact alliinase was stronger (data not presented), suggesting that some proportion of the anti-alliinase IgG may be reactive to the oligosaccharide portion of the alliinase glycoprotein. To determine whether the removal of the oligosaccharide portion was complete, identical membranes were probed with concanavalin A. Concanavalin A showed a strong reaction to intact alliinase, and a faint reaction to the TFMS-treated alliinase. This slight binding was thought to be due to the presence of a residual *N*-acetylglucosamine residue that remains after TFMS treatment (Edge *et al.* 1981), and not to the presence of oligosaccharide side-chains. Deglycosylation reduced the size of the native protein by approximately 3 kDa (M.L. Shaw pers. comm.).

3.4.3. Immunoscreening of an *Allium cepa* cDNA expression library

The *A. cepa* cDNA expression library was screened with I66 anti-alliinase rabbit serum (described in Subsection 2.4.5). Ten plaques were clearly detected by the initial round of screening with I66 anti-alliinase polyclonal sera. Panel A of Figure 3.4 (page 103) shows the signals produced by λAlli4A and λAlli4B. The response from these clones was typical of those selected for further analysis. Two other plaques that showed no reactivity during primary screening (negative controls) were also selected for a further round of screening at low plaque density. Seven of the ten plaques initially selected were again recognized after the second round of screening, while the anti-alliinase sera consistently displayed no affinity to the negative control plaques (Figure 3.4, panels B and C). Secondary screening was carried out at a density of between 50 and 100 plaques per plate which allowed the isolation of individual immunoreactive plaques for further study.

The seven positive clones detected with the polyclonal serum were designated λAlli4A, λAlli4B, λAlli6, λAlli7, λAlli8, λAlli8B and λAlli9. These plaques were isolated and the bacteriophage excised *in vivo* into the phagemid pBluescript (described in Subsection 2.4.6.1). These putative alliinase phagemid clones were designated pAlli4A, pAlli4B, pAlli6, pAlli7, pAlli8, pAlli8B and pAlli9.

Figure 3.3. ***In vitro*** **translation products of** ***A. cepa*** **and positive control haemoglobin poly(A)+ RNA**

In vitro translation reactions were carried out in the presence of ^{35}S-methionine (see Subsection 2.4.4.1.3). Products were separated on 12% SDS-PAGE according to the method of Laemmli (1970) and visualised by exposure to Kodak X-omat AR film.

Lane A. Translation products from *A. cepa* poly(A)+ RNA.

Lane B. Translation products from haemoglobin RNA (positive control).

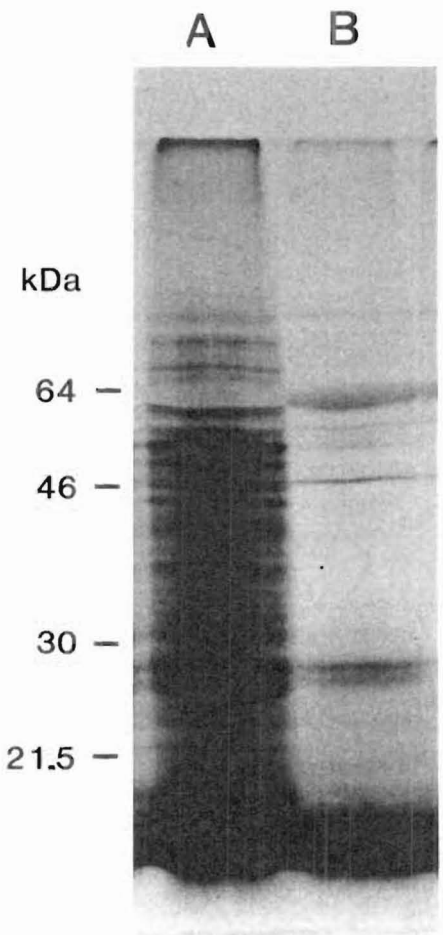

A
B
kDa
64
46
30
21.5

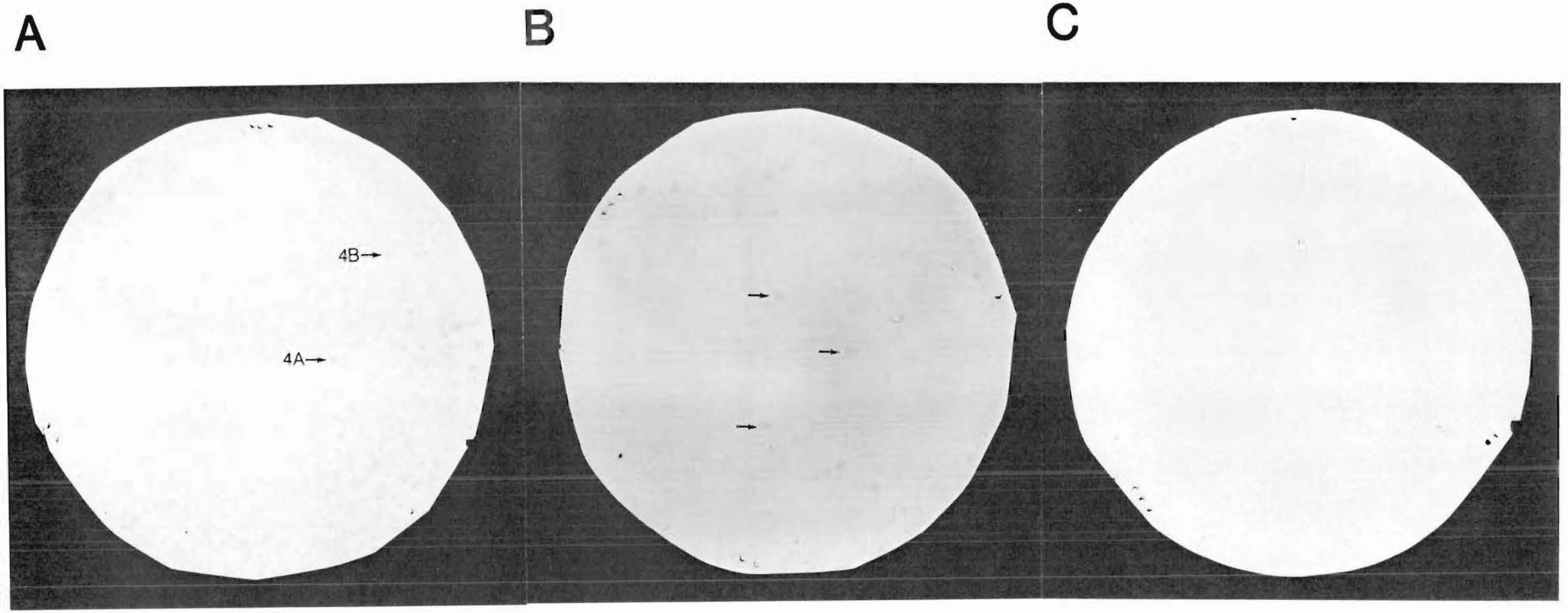
Figure 3.4. Immunoscreening of the *A. cepa* cDNA expression library

An IPTG-impregnated nitrocellulose membrane was overlayed onto a plate containing recombinant bacteriophage and left for 5-6 hours at 37°C to allow the plaques to grow. The membranes were lifted and probed with anti-alliinase antiserum as described in Subsection 2.4.5, page 73. Goat anti-rabbit alkaline phosphatase conjugate was used as a secondary antibody. Alkaline phosphatase activity was assayed using NBT and BCIP.

Panel A. Screening of the *A. cepa* cDNA library was carried out at high density (approximately 2 x 10^4 pfu per plate) and resulted in faint signals. The clones shown are λAlli4A (4A) and λAlli4B (4B).

Panel B. Secondary screening was carried out at lower density (50-100 pfu per plate) and allowed the isolation of pure clones. The clone shown is λAlli4B. Representative positive reactions are arrowed.

Panel C. Secondary screening of plaques which had showed no reactivity to the polyclonal sera during primary screening (negative controls). No signals were detected.


A
B
C
4B
4A

3.5. Analysis of putative alliinase clones

3.5.1. Restriction endonuclease digestion analysis of immunoreactive clones

To initially ascertain the insert size and relationships of the clones, the purified phagemid DNAs were digested with the restriction endonucleases *Xho* I and *Eco* RI (see Subsection 2.4.6.3). This co-digestion would be predicted to excise the entire cDNA insert. This experiment revealed that inserts were present in all the immunopositive clones isolated. Analysis of the resulting fragment sizes showed that the cDNAs Alli4B, Alli6, Alli7, Alli8, and Alli8B ranged in size from 1.6 kb to 1.7 kb, and all contained an *Eco* RI site approximately 200 bp from the 5' end (Figure 3.5, page 115). The Alli4A cDNA did not have sites for either of these enzymes and was estimated to be approximately 0.8 kb long. Alli4B and Alli8 contained an *Xho* I site near the 5' end which was not present in the other clones. The Alli9 sequence was about 2.4 kb in length and contained an *Xho* I site approximately 0.7 kb from the 3' end (Figure 3.5). The isolation of these putative alliinase clones using anti-alliinase antisera and the similarity in their restriction patterns (except for Alli4A and Alli4B) indicated a high level of relatedness. Consequently further characterization was carried out by DNA sequence analysis.

3.5.2. Sequence determination of the presumptive alliinase clones and computer-aided DNA sequence analysis

The nucleotide sequence of the cDNA inserts of pAlli4A, pAlli4B, pAlli6, pAlli7, pAlli8B and pAlli9 were determined within the pBluescript phagemid by double-stranded techniques (see Subsection 2.4.6.4) using the dideoxy-chain termination method (Sanger *et al.* 1977). The clone pAlli8 was not sequenced as it appeared to be identical in the location of its *Xho* I and *Eco* RI restriction sites to pAlli4B. Restriction mapping had suggested that Alli6 and Alli8 were also identical, but as this similarity was based on the location of a single *Eco* RI site it was decided sequence analysis would be valuable in clarifying the relationship of these two clones. The unmodified pBluescript vector contains universal M13/pUC forward and reverse and KS/SK sequencing primer annealing sites at the borders of the multicloning site (Short *et al.* 1988). However cloning of insert DNA into the *Xho* I site of the λZAPII vector destroys the KS primer site, therefore initial sequencing from the two vector/insert boundaries was carried out with the pUC/M13 universal

forward primer and the SK primer (described in Subsection 2.4.6.4). Phagemid DNA purified by hydroxyapatite column chromatography was found to be of adequate purity to result in clearly readable sequences (Figure 3.6, page 117). Preliminary sequence data from the vector-insert borders indicated a strong sequence similarity between the cDNA inserts of pAlli4B, 6, 7, 8B and the 5' end of Alli9, although the clone pAlli4A contained an unrelated DNA sequence. Therefore sequencing of the cDNAs proceeded by using 17 nucleotide-long primers which were designed from the DNA sequences as they were determined (shown in Table 2.3). Figure 3.7 on page 119 shows the strategy used to sequence the clones.

The DNA sequences of the putative alliinase clones were aligned using several available sequence analysis computer packages (see Subsection 2.4.6.5). The sequences of Alli4B, 6, 7, 8B and 9 were mostly identical and the composite sequence is presented in Figure 3.8 (page 121). Alli7 was the longest of these analogous sequences and all positions in Figure 3.8 are numbered relative to the start of this cDNA. Two similar clones, Alli4B and Alli6, were sequenced along their entire length, and the other related clones were partially sequenced. Alli7 was sequenced from the 5' end (base 1) to position 723, and from position 1220 to the 3' end (base 1757). Alli8B was sequenced from the 5' end to position 735, and from position 1145 to the 3' end. Alli9 was sequenced from its 5' end to position 1186, and from position 1591 to the 3' end which was 630 bases further than those of the other homologous clones. This resulted in each part of the overlapping sequence being read from different cDNA clones at least three times as shown in Figure 3.7. Alli4A was completely sequenced in both directions. Within the overlapping regions sequenced, the cDNAs Alli6, Alli7, Alli9 and Alli8B varied only in the length of their 5' leader regions. Alli4B was identical to Alli6 throughout but had 37 bases less at the 3' end and 6 nonhomologous bases just prior to the polyadenylate tail (Figure 3.8). The 5' end of the cDNA Alli7 was 111 bases longer than any of the other clones.

The polyadenylate tails of all the clones were found to be 18 nucleotides long, which is equal to the number of dTTP residues present in the *Xho* I-$(dT)_{18}$ cDNA synthesis primer. Three potential polyadenylation signals were present among these homologous sequences. Two overlapping signals were present between nucleotides 1669 and 1678, and a third signal began at position 1719 (Figure 3.8). Alli4B did not contain the polyadenylation signal beginning at position 1719.

A computer-generated restriction map of this composite sequence is presented in Figure 3.9 (page 123). An *Eco* RI recognition site was found at position 328. This confirmed the evidence presented in Figure 3.5, where digestion of the putative alliinase clones with *Eco* RI and *Xho* I resulted in the resolution of two fragments

from these cDNA inserts. This restriction enzyme analysis had also indicated that the cDNA insert of pAlli9, at approximately 2.4 kb, was much longer than the other clones isolated (Figure 3.5). DNA sequence analysis showed that the majority of the cDNA contained within pAlli9 was identical to Alli6, except that it was four bases shorter at the 5'-end, and that the longer insert size was due to the presence at the 3' end of a further 630 bp of cDNA which was unrelated in sequence to the other clones. As these two cDNA sequences were found to be connected by an undigested *Xho* I-$(dT)_{18}$ linker region, it is probable that they were cloned in tandem as a result of incomplete *Xho* I digestion during library construction. This extraneous 630 bp cDNA sequence is presented in Appendix I. DNA sequencing further revealed that the *Xho* I site at the 5'-end of Alli6 resulted from the artifactual ligation of an *Xho* I-$(dT)_{18}$ linker-primer and was not due to the presence of an *Xho* I site within this cDNA sequence. The unsequenced clone, pAlli8, also had an *Xho* I site at the 5' end indicating that it was likely to be identical to Alli6.

Although the anti-alliinase antiserum I66 showed high affinity to the protein expressed by the pAlli4A clone, DNA sequencing revealed that no similarity existed between the Alli4A cDNA and the other cDNAs isolated. The Alli4A cDNA insert was 826 bp long including a 18 nucleotide-long poly(A) tail. It contained a longest ORF which started at a methionine codon at position 91 and ended at a stop codon at position 697. No upstream stop codons were present, therefore the possibility that this sequence encodes a partial reading frame cannot be discounted. This 216 amino acid-deduced peptide sequence has a predicted molecular weight of 24,207. This sequence is presented in Appendix II.

3.5.2.1. Identification of ORFs among the related cDNA sequences

The composite sequence of Alli4B, Alli6, Alli7, Alli8B and Alli9 was searched for ORFs. The resultant predicted amino acid sequences were compared to the amino acid sequence data of intact and cyanogen bromide-generated fragments of alliinase purified from the *A. cepa* cultivar Pukekohe Longkeeper (presented in Table 3.1). One reading frame encoded amino acid sequences almost identical to those directly determined from the alliinase protein. This reading frame was the longest found, beginning at a methionine codon at position 136 and terminating at a stop codon at position 1573. A second in-frame stop codon was present at position 1609. The N-terminal 23 amino acid-peptide sequence, ALN1, aligned with the predicted protein sequence beginning at the 35th amino acid of the longest ORF (Figure 3.8). Residues 11, 20 and 23 of ALN1 differed from those predicted by the cDNA

sequence. The three peptide sequences determined from cyanogen bromide-cleaved fragments of alliinase also showed high homology to the predicted peptide sequence from the same ORF. ALN3 (Table 3.1) was encoded by nucleotides 511 to 549, ALN4 was encoded by nucleotides 916 to 966, and ALN5 was encoded by nucleotides 1483 to 1517. An alanine residue identified within ALN3 aligned with a serine residue predicted by the cDNA sequence. ALN2 (shown in Table 3.1) did not align to the protein sequences predicted from the cDNA clones. This peptide sequence showed high homology to the lectins of *A. sativum* (see Subsection 3.1) and therefore was probably obtained from an *A. cepa* lectin contaminant of the alliinase preparation. Translation of the two other cDNA sequence possible reading frames did not reveal any peptide sequences homologous to those determined from alliinase protein.

This strong correspondence between the amino acid sequence determined from purified alliinase with that derived from the cDNA clones is evidence that these cDNAs are derived from alliinase transcripts. A single ORF encoded all of these peptide sequences, and was therefore taken to be the most likely correct reading frame.

The cDNAs Alli8B, 6, 4B and 9 contained between 24 and 16 nucleotides upstream of the start of the longest ORF. The cDNA sequence Alli7 was longer, with 135 nucleotides upstream of the longest ORF. This long 5' leader sequence contained multiple potential translation initiation and termination codons. Methionine codons in the same phase as the longest ORF were located at positions 10, 46 and 61, and were followed by in-frame stop codons at positions 103, 118 and 127. Also present were two methionine codons in a different reading frame, at positions 17 and 71, which were followed by stop codons in the same frame at positions 20 and 131. A further stop codon was present in the third ORF of this upstream region, at position 54 (Figure 3.8).

3.5.2.2. Base composition and codon usage

The guanine and cytosine (G+C) content of the putative alliinase cDNA sequences showed little variation overall (39.9% to 40%). Within the identical longest ORF the G+C content was 41%. The longer 5' region of Alli7 contained 37% G+C as compared to 43% within the shorter 5' regions of the other homologous clones. The G+C content of the shorter 3' untranslated region of Alli4b was 32%, as compared to 29% for that of the other alliinase clones.

A codon usage profile for alliinase is shown in Table 3.5. Within the longest ORF, codon usage analysis indicated a predominantly higher percentage adenine and

thymidine (A+T) than G+C content in the degenerate third base, with 28.2%, 31.1%, 21.1% and 19.6% of the codons terminating in A, T, G or C respectively. Of the 61 amino acid-encoding codons, 58 are used to encode alliinase.

Table 3.5. Codon usage of the 479-residue alliinase precursor derived by translation of the cDNA sequence longest open reading frame

Base: 1st	2nd								3rd
	A		C		G		T		
A	21	Lys	11	Thr	7	Arg	9	Ile	A
A	8	Asn	2	Thr	3	Ser	10	Ile	C
A	17	Lys	2	Thr	4	Arg	15	Met	G
A	19	Asn	13	Thr	8	Ser	9	Ile	T
C	10	Gln	10	Pro	3	Arg	6	Leu	A
C	4	His	0	Pro	2	Arg	5	Leu	C
C	8	Gln	1	Pro	0	Arg	4	Leu	G
C	6	His	8	Pro	4	Arg	8	Leu	T
G	21	Glu	14	Ala	11	Gly	3	Val	A
G	7	Asp	4	Ala	4	Gly	2	Val	C
G	10	Glu	2	Ala	6	Gly	10	Val	G
G	14	Asp	7	Ala	8	Gly	11	Val	T
T	1	Ter	8	Ser	0	Ter	0	Leu	A
T	18	Tyr	8	Ser	9	Cys	8	Phe	C
T	0	Ter	1	Ser	9	Trp	11	Leu	G
T	8	Tyr	10	Ser	3	Cys	13	Phe	T

3.5.2.3. Amino acid sequence analysis

The protein encoded by the longest ORF of the alliinase cDNA sequence is 479 amino acids long with a predicted molecular mass of 54 884 Da. Hydropathy analysis (Figure 3.10, Panel A. page 128) of the inferred alliinase amino acid sequence by the method of Kyte & Doolittle (1982) predicts that the mature alliinase polypeptide is predominantly hydrophilic. The regions between residues 160 to 162, and 327 to 330 have an average hydrophobicity exceeding 1.5. The regions between residues 101 to 108, 197 to 216 and 309 to 321 were predicted to show a markedly hydrophilic nature. The 34-residue N-terminal region contained a core of highly hydrophobic residues, indicative of a signal peptide (Davis & Tai 1980). A putative signal peptidase recognition sequence of Ala-X_{aa}-Ala[8] (Perlman & Halvorsen 1983), is present at amino acid positions 32 to 34, immediately preceding the site of alignment of the mature alliinase N-terminal peptide sequence. Cleavage between

[8]where X_{aa} is any amino acid

the alanine residue at position 34 and the lysine at position 35 would produce an inferred 445 amino acid mature alliinase protein with a predicted molecular weight, in the absence of post-translational modifications, of 51,135 Da.

A 'weight matrix' approach for the prediction of signal sequence cleavage sites (von Heijne 1986) was used to assess the most probable length of the predicted alliinase signal peptide. This method is based on a probability matrix constructed from an analysis of the residues flanking 161 known eukaryotic signal sequence cleavage sites. The protein sequence of interest is scanned by a moving window of a set number of residues. Respective probabilities are allocated from the weight matrix for each amino acid in the window dependent on its location, and the sum of these values is then assigned to the central residue. This method thereby measures the fit of the sequence contained within the window to those used to construct the weight matrix. The highest scoring position based on this approach has been shown to correctly identify 125 out of 161 eukaryotic cleavage sites (von Heijne 1986). Applying this approach to the predicted N-terminus of the alliinase cDNA longest ORF translation product resulted in the valine residue at position 27 having the highest score (1.16), and the alanine residue at position 34 having the second highest score (0.67).

An analysis of the antigenicity of the predicted alliinase peptide was performed according to the method of Welling *et al.* (1985). This approach is based on the abundance of each amino acid in known antigenic determinants compared with the average protein composition. Multiple antigenic sites are predicted to lie throughout the alliinase peptide with the two most antigenic regions residing between positions 151 to 154 and positions 188 to 195 (Figure 3.10, Panel B). It is not known whether either of these regions are highly immunogenic.

A comparison of the amino acid sequence composition of the inferred alliinase peptide (beginning at amino acid 35) with that previously determined by acid hydrolysis of alliinase protein isolated from a yellow-skinned cultivar of *A. cepa* (Nock & Mazelis 1987) is shown in Table 3.6 (page 110). The estimated content of aspartic acid and histidine predicted from the cDNA-derived sequence was markedly lower, whereas the estimated amounts of cysteine and methionine were higher. The proportions of asparagine and glutamine were not provided by Nock & Mazelis (1987), and it is therefore conceivable that these values have been incorporated into those given for aspartic acid and glutamic acid respectively, which would be more consistent with the proportions calculated from the cDNA-derived alliinase protein. The reasons for the differences in the abundance of histidine, cysteine and methionine are not clear, but may be a reflection of the different cultivar used in this analysis. Estimates for the other amino acids were in reasonable agreement.

Table 3.6. Comparison of the *A. cepa* alliinase amino acid composition derived from cDNA clones with that determined from purified alliinase

	Values calculated from the deduced amino acid sequence[§]			Direct analysis[a]
	Number	Weight %	Moles %[¶]	Moles %[¶]
Alanine	25	3.5	5.6	6.0
Cysteine	11	2.2	2.5	0.4*
Aspartic acid	21	4.7	4.7	11.1*
Glutamic acid	31	7.8	7.0	9.5
Phenylalanine	20	5.8	4.5	4.4
Glycine	27	3.0	6.1	8.2
Histidine	9	2.4	2.0	7.0*
Isoleucine	23	5.1	5.2	4.8
Lysine	36	9.0	8.1	6.0
Leucine	31	6.9	7.0	6.9
Methionine	12	3.1	2.7	0.6*
Asparagine	23	5.1	5.2	N.D
Proline	18	3.4	4.0	4.2
Glutamine	17	4.3	3.8	N.D
Arginine	21	6.4	4.7	4.4
Serine	33	5.6	7.4	7.7
Threonine	28	5.5	6.3	5.5
Valine	25	4.9	5.6	6.9
Tryptophan	9	3.3	2.0	1.2
Tyrosine	25	8.0	5.6	5.2

§ Comparisons are based on the deduced peptide lacking the N-terminal 34 amino acids.

[a] Values determined by acid hydrolysis from Nock & Mazelis (1987). The cultivar used was an unspecified yellow bulb variety.

¶ Percentages are based on the mol amino acid per mol protein values.

* Estimates that vary markedly between the two approaches.

N.D. Not determined.

Comparison of the derived amino acid composition from that of average values for eukaryotes (Doolittle 1987) did not indicate any amino acid was significantly under- or over-represented. The calculated isoelectric point of the inferred protein lacking the 34 amino acid putative signal peptide region is 8.18, which is similar to a pI of 8.0 estimated previously for *A. cepa* alliinase from isoelectric focusing studies (Nock & Mazelis 1987), although the possibility that post-translational modifications *in vivo* may affect the pI value cannot be ignored.

Examination of the predicted alliinase protein sequence for conserved sequence motifs shows that the four asparagine residues at positions 53, 180, 225 and 362 lie within the Asn-linked glycosylation site sequence motif Asn-X_{aa}-Ser/Thr (Kornfeld & Kornfeld 1985). Four possible PLP attachment motifs, Ser/Thr-X_{aa}-X_{aa}-Lys (Tanase *et al.* 1979), are present within the derived alliinase amino acid sequence at positions 271, 285, 377 and 440 (Figure 3.8). Residues flanking these sites did not show any resemblance to those surrounding previously identified PLP-binding sites compiled within the PROSITE database (Bairoch 1992).

Secondary structure predictions of the derived peptide were made using the methods of Chou & Fasman (1974), as well as Garnier *et al.* (1978) with the PeptideStructure and PlotStructure programs of the UWGCG package. The Garnier *et al.* (1978) prediction is presented in Figure 3.11 on page 130. The data for generating both predictions is presented in Appendix III. Eight regions are predicted by both methods to be α-helical (Figure 3.11). The largest of these spans the 14 residues beginning at position 36 and aligns with the predicted N-terminus of the mature peptide. Seven regions of β-strand and sixteen turns are also predicted by both methods used. This results in several regions of β-turns (β-strands linked by reverse turns) distributed throughout the peptide, the largest of which is located between residues 53 and 100. Five cysteine residues predicted from the longest ORF translation product are clustered between amino acids 74 and 91 and lie within this largest region of predicted β-turns. Two predicted cysteine residues are in close proximity at positions 405 and 413, while other predicted cysteine residues are more broadly distributed throughout the sequence. According to both predictions (Figure 3.11), the two potential Asn-linked glycosylation sites at residues 53 and 180 are within regions of β-turns, whereas the two sites toward the C-terminus are in regions of random coil.

3.5.2.4. Comparison of the cDNA sequences to databases

The program FASTA (Pearson & Lipman 1988) was used to determine sequence homology at the nucleotide level between alliinase and entries in the EMBL (Release 31) and GenBank (Release 72) databases (see Subsection 2.4.6.5). The method employed by FASTA generated best fits by inserting gaps into the sequence and allowing mismatches as required (Pearson & Lipman 1988). There were no significant homologies found between the *A. cepa* alliinase sequence and any database entry. The program BLASTP (Altschul *et al.* 1990) was used to determine sequence homology at the amino acid sequence level between alliinase and entries in the SwissProt (release 22) and GenPept databases (release 30). The BLASTP program uses a heuristic search algorithm which first identifies locally optimal sequence alignments and then scores the aligned identical and differing residues in those regions using an amino acid replaceability matrix (Schwartz & Dayhoff 1978). The greatest match found, at 10%, was less than the 11% calculated to occur within the database by chance alone. Therefore the slight homologies found between regions of the *A. cepa* alliinase translation product and entries in the protein sequence databases were not considered to be biologically significant.

The non-alliinase cDNA clone, Alli4A, and the 630 bp sequence present at the 3' end of the Alli9 cDNA, were also compared to the sequences contained within the GenBank DNA and the SwissProt protein databases. These searches revealed a high identity between the Alli9 630 bp tail sequence and the cyclophilins (peptidyl-prolyl *cis-trans* isomerases). The greatest correspondence was to the cyclophilins of *Brassica napus* and *Lycopersicon esculentum*, with the predicted protein sequence showing more than 85% homology to that of the *L. esculentum* cyclophilin. Neither the DNA sequence nor the predicted amino acid sequence of the Alli4A clone showed any similarity to the DNA or protein sequences available in the databases nor to the other clones. No homologies between any of these sequences and those of various vector sequences available through the GenBank database were found in either orientation.

3.5.3. Expression of Alli6 recombinant protein in *E. coli*

The alliinase cDNA clones were contained within the IPTG-inducible phagemid vector pBluescript(-) due to their *in vivo* excision from the λZAPII expression vector (see Subsection 2.4.6.1). Expression of the encoded polypeptide was induced by growing *E. coli* XL1-Blue containing the phagemid pAlli6 (XLalli6) in the presence

of IPTG (described in Subsection 2.4.6.6). Total proteins were extracted in Laemmli sample buffer from IPTG-induced and uninduced XLalli6, as well as from non-recombinant pBluescript (XL(-)), and separated by SDS-PAGE (Subsection 2.8.2). The extracted proteins were examined by silver staining and immunoblotting with anti-alliinase polyclonal antiserum, as described in Subsection 2.4.6.6 and shown in Figure 3.12 on page 132.

Silver staining of SDS-PAGE gels demonstrated that band intensities in each lane were comparable (Figure 3.12, Panel A). Western blotting of these gels and probing with anti-alliinase antiserum recognized two polypeptides, a major band of 47 kDa and another of 41.4 kDa. Neither of these immunoreactive bands were detected among the proteins extracted from XL(-) (Figure 3.12, Panel B, Lane 3). These data may be compared with a molecular mass of 47 kDa estimated for deglycosylated mature alliinase by SDS-PAGE (M.L. Shaw pers. comm.) and 54.9 kDa predicted by back-translation of the 479 amino acid longest ORF. Both of these immunoreactive proteins were expressed by XLalli6 either with or without induction by IPTG (Figure 3.12, Panel B, Lane 1 and Lane 2). This is most likely to result from the use of yeast extract in the culture media. Yeast extract contains forms of the *lac* activator capable of inducing expression of a cloned gene in the absence of IPTG (Duffaud *et al.* 1987). The 47 kDa protein did not appear to be highly over-expressed, as there was no discernibly stronger band of this size among the silver-stained protein profiles. Furthermore, its presence appeared to be masked by the co-migration of endogenous *E. coli* proteins. The 41.4 kDa peptide was visible among the silver-stained protein profiles of these clones (Figure 3.12, Panel A, arrowed).

3.6. Screening of an *A. cepa* genomic library with a putative alliinase cDNA clone

The *A. cepa* genomic clone bank (construction described in Materials & Methods Section 2.2) was probed with a 1.4 kb *Xho* I/*Eco* RI fragment excised from pAlli9. Approximately 2.6 x 10^5 pfu from the library were screened and 23 putative positive plaques were selected by alignment of duplicate filters (Subsection 2.5). None of these bacteriophage consistently produced positive hybridization signals on secondary screening, therefore no further genomic clone isolation studies were performed.

Figure 3.5. Restriction map derived from restriction endonuclease analysis of immuno-selected clones

Phagemid DNA from each clone was digested with *Eco* RI and *Xho* I endonucleases individually or in combination (see Subsection 2.4.6.3). The numbers denote the cDNAs Alli4A, Alli4B, Alli6, Alli7, Alli8, Alli8B and Alli9.

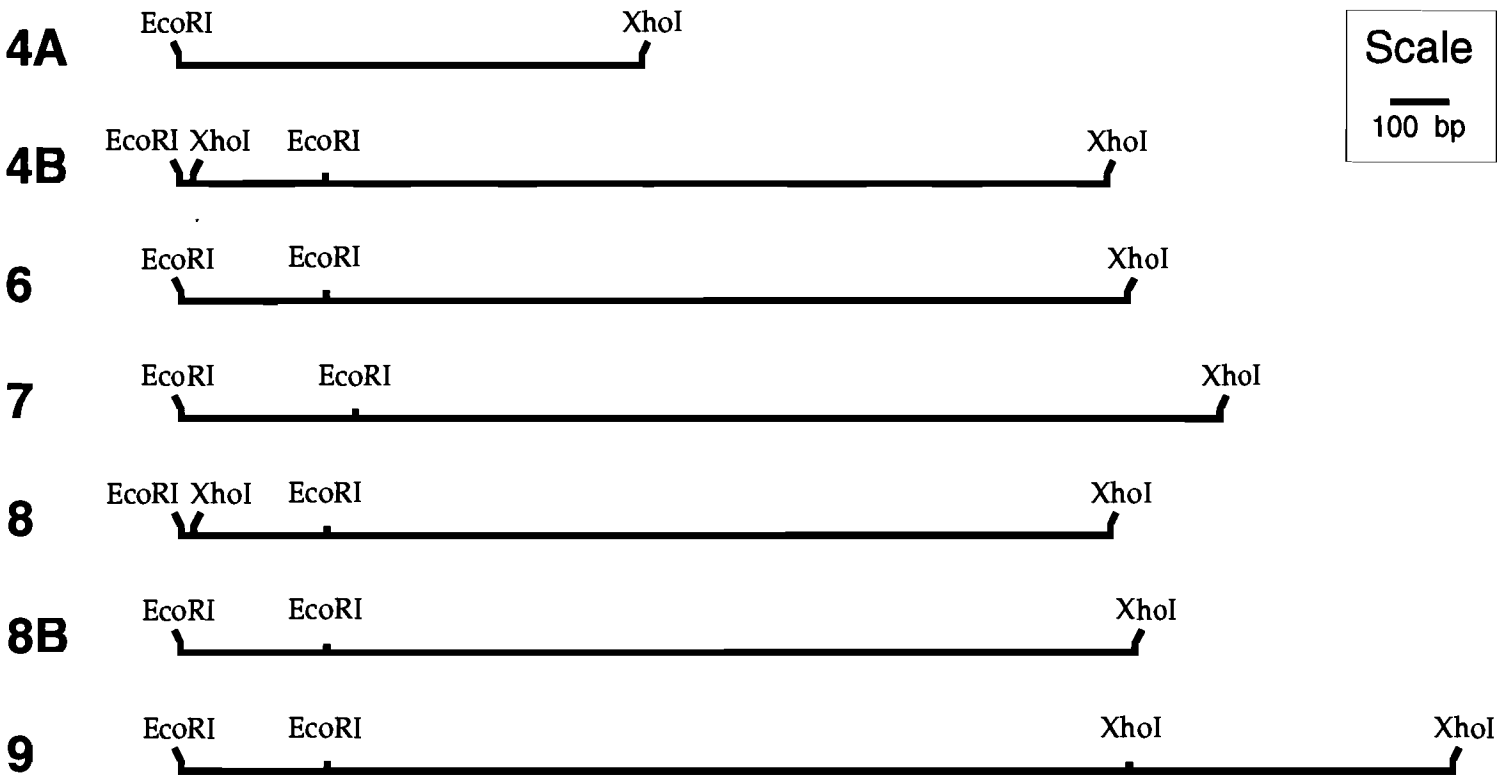

4A
EcoRI
XhoI
4B
EcoRI
XhoI
EcoRI
XhoI
6
EcoRI
EcoRI
XhoI
7
EcoRI
EcoRI
XhoI
8
EcoRI
XhoI
EcoRI
XhoI
8B
EcoRI
EcoRI
XhoI
9
EcoRI
EcoRI
XhoI
XhoI
Scale
100 bp

Figure 3.6. Nucleotide sequencing of putative alliinase cDNA clones

The autoradiograph displaying the DNA sequence of the Alli6 cDNA is presented between positions 700 and 770 (see Subsection 2.4.6.4 for method). The region containing the sequence of the 17mer sequencing primer SC6 is bracketed.

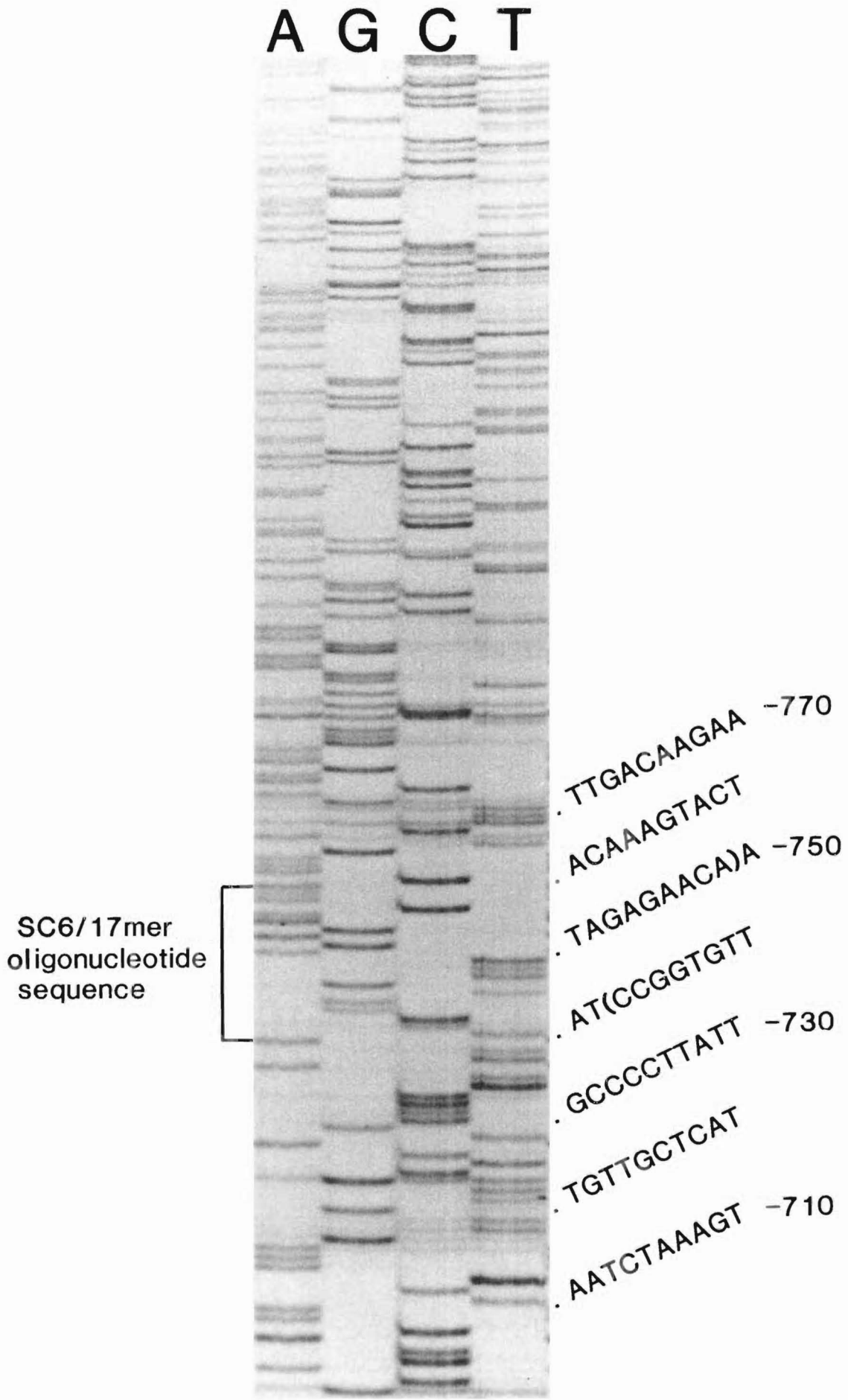

A G C T
SC6/17mer
oligonucleotide
sequence
. TTGACAAGAA -770
. ACAAAGTACT
. TAGAGAACA)A -750
. AT(CCGGTGTT
. GCCCCTTATT -730
. TGTTGCTCAT
. AATCTAAAGT -710

Figure 3.7. Sequencing schedule of the presumptive alliinase cDNA clones

The presumptive alliinase clones were excised from λZAPII *in vivo* in the phagemid pBluescript(-) and sequenced by the dideoxy-chain termination method from double-stranded phagemid DNA (see Subsection 2.4.6.4). The universal pUC/M13 forward and SK sequencing primers were used to sequence across the vector-insert borders. Further 17mer oligonucleotide primers, SC1, SC2, SC3, SC4, SC5, SC6, SC7, SC8 and SC9 were designed from the DNA sequence as it was determined (shown in Table 2.3). The arrows indicate the direction and extent of the sequence determination.

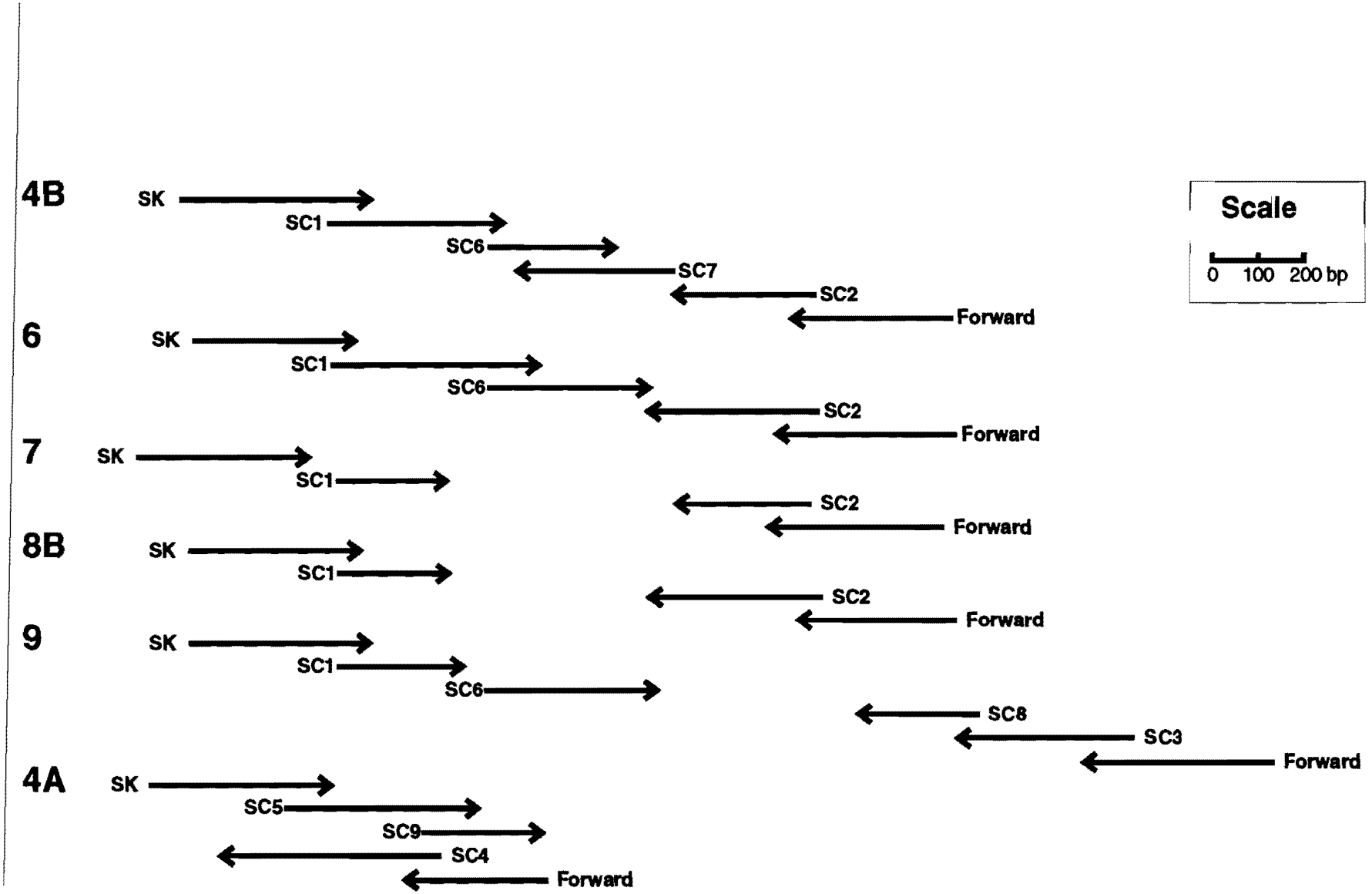

4B
SK
SC1
SC6
SC7
SC2
Forward
6
SK
SC1
SC6
SC2
Forward
7
SK
SC1
SC2
Forward
8B
SK
SC1
SC2
Forward
9
SK
SC1
SC6
SC8
SC3
Forward
4A
SK
SC5
SC9
SC4
Forward
Scale
0 100 200 bp

Figure 3.8. Nucleotide and derived amino acid sequence of the putative alliinase cDNAs Alli4B, Alli6, Alli7, Alli8B and Alli9

Numbering is relative to the longest cDNA sequence, Alli7. The 5' ends of the other sequences are marked by a •, and lie between position 112 and 120. The truncated 3'-end of the cDNA Alli4B is located underneath the sequence from position 1696. The amino acid sequence derived from the longest open reading frame is shown above the DNA sequence. The standard one-letter code is used and the amino acid sequence is numbered with 1 being the methionine of the longest open reading frame.

Key to symbols.

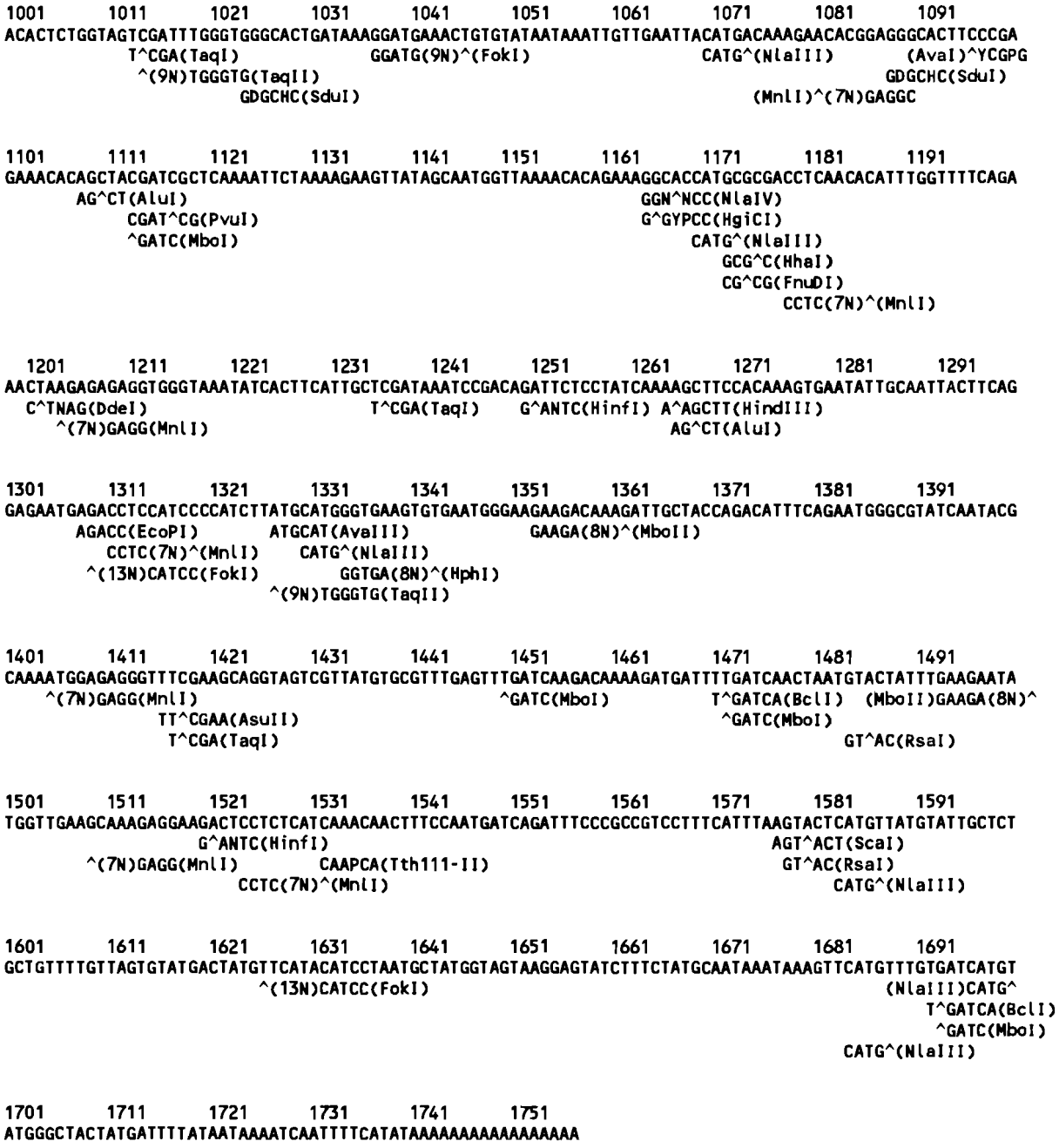
XXX - predicted amino acid sequence aligning with that determined from pure alliinase protein.
ATG - possible translation initiation start codons.
TAA, TAG - possible translation termination codons.
AATAAA - potential polyadenylation signals.
⌐¬ - potential Asn-linked glycosylation sites.
⌐¬ - potential PLP binding sites.

```
                •7
1       TTT GGT TTG ATG AAA TAT GTA GGC TGC CTC ACT TTC TAT TTA CCC ATG TCT GTT AAT  57

                                                                          •8B
58      TCC ATG TTG CGG AAT GGA ACA TAC CTC ATT TCA AAA CAT TTT CAG TAA GCG AGA TAT  114

          •6/4B •9                        M   E   S   Y   H   K   V   G   S   N   K   M 12
115     CCA TAG CAG AGC TAA TTA GCT ATG GAG TCT TAC CAC AAA GTT GGC AGT AAT AAA ATG  171

   13   P   S   L   L   I   L   I   C   I   I   M   S   S   F   V   N   N   N   I  31
172     CCA AGC CTT CTT ATT TTG ATA TGC ATA ATC ATG TCT TCA TTT GTT AAC AAT AAT ATA  228

   32   A   Q   A * K   V   T   W   S   L   K   A   A   E   E   A   E   A   V   A  50
229     GCT CAA GCG AAG GTG ACA TGG AGT TTG AAG GCA GCA GAA GAG GCA GAA GCA GTG GCT  285

   51   N   I   N   C   S   G   H   G   R   A   F   L   D   G   I   L   S   D   G  69
286     AAT ATA AAC TGT TCA GGG CAT GGG AGA GCT TTT TTG GAC GGA ATT CTT TCA GAT GGA  342

   70   S   P   K   C   E   C   N   T   C   Y   T   G   A   D   C   S   E   K   I  88
343     TCT CCT AAA TGC GAG TGC AAT ACT TGC TAC ACT GGT GCA GAC TGC TCT GAA AAG ATT  399

   89   T   G   C   S   A   D   V   A   S   G   D   G   L   F   L   E   E   Y   W  107
400     ACA GGT TGC TCT GCT GAT GTT GCC AGT GGT GAT GGA CTG TTT CTA GAA GAA TAC TGG  456

  108   Q   Q   H   K   E   N   S   A   V   L   V   S   G   W   H   R   M   S   Y  126
457     CAG CAG CAC AAG GAA AAC AGT GCA GTG CTG GTT TCA GGA TGG CAC AGA ATG AGC TAC  513

  127   F   F   N   P   V   S   N   F   I   S   F   E   L   E   K   T   I   K   E  145
514     TTT TTC AAC CCA GTT AGC AAT TTC ATA TCT TTC GAG CTT GAA AAA ACA ATT AAA GAA  570

  146   L   H   E   I   V   G   N   A   A   A   K   D   R   Y   I   V   F   G   V  164
571     CTA CAT GAG ATA GTC GGA AAT GCT GCT GCA AAG GAC AGG TAC ATT GTG TTT GGA GTA  627

  165   G   V   T   Q   L   I   H   G   L   V   I   S   L   S   P   N   M   T   A  183
628     GGG GTG ACT CAA CTC ATC CAT GGA TTG GTC ATC TCC CTT TCA CCA AAT ATG ACT GCC  687

  184   T   P   C   A   P   Q   S   K   V   V   A   H   A   P   Y   Y   P   V   F  202
688     ACT CCT TGT GCA CCA CAA TCT AAA GTT GTT GCT CAT GCC CCT TAT TAT CCG GTG TTT  741

  203   R   E   Q   T   K   Y   F   D   K   K   G   Y   E   W   K   G   N   A   A  221
742     AGA GAA CAA ACA AAG TAC TTT GAC AAG AAA GGG TAC GAG TGG AAA GGA AAT GCA GCG  798

  222   D   Y   V   N   T   S   T   P   E   Q   F   I   E   M   V   T   S   P   N  240
799     GAT TAC GTG AAC ACT TCA ACT CCA GAG CAA TTC ATT GAG ATG GTT ACT TCA CCT AAT  855

  241   N   P   E   G   L   L   R   H   E   V   I   K   G   C   K   S   I   Y   D  259
856     AAC CCA GAA GGT CTG CTT CGC CAT GAA GTA ATA AAG GGA TGC AAA TCC ATC TAC GAT  912

  260   M   V   Y   Y   W   P   H   Y   T   P   I   K   Y   K   A   D   E   D   I  278
913     ATG GTT TAC TAC TGG CCT CAT TAC ACC CCA ATC AAG TAC AAA GCC GAT GAA GAT ATC  969

  279   M   L   F   T   M   S   K   Y   T   G   H   S   G   S   R   F   G   W   A  297
970     ATG TTG TTT ACA ATG TCT AAA TAC ACT GGA CAC TCT GGT AGT CGA TTT GGG TGG GCA 1026

  298   L   I   K   D   E   T   V   Y   N   K   L   L   N   Y   M   T   K   N   T  316
1027    CTG ATA AAG GAT GAA ACT GTG TAT AAT AAA TTG TTG AAT TAC ATG ACA AAG AAC ACG 1083

  317   E   G   T   S   R   E   T   Q   L   R   S   L   K   I   L   K   E   V   I  335
1084    GAG GGC ACT TCC CGA GAA ACA CAG CTA CGA TCG CTC AAA ATT CTA AAA GAA GTT ATA 1140

  336   A   M   V   K   T   Q   K   G   T   M   R   D   L   N   T   F   G   F   Q  354
1141    GCA ATG GTT AAA ACA CAG AAA GGC ACC ATG CGC GAC CTC AAC ACA TTT GGT TTT CAG 1197

  355   K   L   R   E   R   W   V   N   I   T   S   L   L   D   K   S   D   R   F  373
1198    AAA CTA AGA GAG AGG TGG GTA AAT ATC ACT TCA TTG CTC GAT AAA TCC GAC AGA TTC 1254

  374   S   Y   Q   K   L   P   Q   S   E   Y   C   N   Y   F   R   R   M   R   P  392
1255    TCC TAT CAA AAG CTT CCA CAA AGT GAA TAT TGC AAT TAC TTC AGG AGA ATG AGA CCT 1311

  393   P   S   P   S   Y   A   W   V   K   C   E   W   E   E   D   K   D   C   Y  411
1132    CCA TCC CCA TCT TAT GCA TGG GTG AAG TGT GAA TGG GAA GAA GAC AAA GAT TGC TAC 1368

  412   Q   T   F   Q   N   G   R   I   N   T   Q   N   G   E   G   F   E   A   G  430
1369    CAG ACA TTT CAG AAT GGG CGT ATC AAT ACG CAA AAT GGA GAG GGT TTC GAA GCA GGT 1425

  431   S   R   Y   V   R   L   S   L   I   K   T   K   D   D   F   D   Q   L   M  449
1426    AGT CGT TAT GTG CGT TTG AGT TTG ATC AAG ACA AAA GAT GAT TTT GAT CAA CTA ATG 1482

  450   Y   Y   L   K   N   M   V   E   A   K   R   K   T   P   L   I   K   Q   L  468
1483    TAC TAT TTG AAG AAT ATG GTT GAA GCA AAG AGG AAG ACT CCT CTC ATC AAA CAA CTT 1539

  469   S   N   D   Q   I   S   R   R   P   F   I
1540    TCC AAT GAT CAG ATT TCC CGC CGT CCT TTC ATT TAA GTA CTC ATG TTA TGT ATT GCT 1596

1597    CTG CTG TTT TGT TAG TGT ATG ACT ATG TTC ATA CAT CCT AAT GCT ATG GTA GTA AGG 1653

1654    AGT ATC TTT CTA TGC AAT AAA TAA AGT TCA TGT TTG TGA TCA TGT ATG GGC TAC TAT 1710
4B                                                          AAG GAT AAA AAA AAA

1711    GAT TTT ATA ATA AAA TCA ATT TTC ATA TAA AAA AAA AAA AAA AAA A   1757
4B      AAA AAA AAA
```

Figure 3.9. Computer generated restriction map of the *A. cepa* alliinase cDNA composite sequence

A restriction map of the cDNA sequence was generated using the program DM. The first base of the recognition sequence is aligned to the cDNA sequence and the position of cleavage is indicated by a ^. Restriction endonucleases without a recognition sequence are listed.

```
       11        21        31        41        51        61        71        81        91
TTTGGTTTGATGAAATATGTAGGCTGCCTCACTTTCTATTTACCCATGTCTGTTAATTCCATGTTGCGGAATGGAACATACCTCATTTCAAAACATTT
                          CTC(7N)^(MnlI)                    CATG^(NlaIII)       CCTC(7N)^(MnlI)
               ^(12N)GCTGC(BbvI)          CATG^(NlaIII)
                    GC^NGC(Fnu4HI)

101       111       121       131       141       151       161       171       181       191
TCAGTAAGCGAGATATCCATAGCAGAGCTAATTAGCTATGGAGTCTTACCACAAAGTTGGCAGTAATAAAATGCCAAGCCTTCTTATTTTGATATGCATA
          GAT^ATC(EcoRV)       AG^CT(AluI)    CCA(5N)^NTGG(BstXI)                          ATGCAT(AvaIII)
                      AG^CT(AluI)    G^ANTC(HinfI)

201       211       221       231       241       251       261       271       281       291
ATCATGTCTTCATTTGTTAACAATAATATAGCTCAAGCGAAGGTGACATGGAGTTTGAAGGCAGCAGAAGAGGCAGAAGCAGTGGCTAATATAAACTGTT
  CATG^(NlaIII)             AG^CT(AluI) GGTGA(8N)^(HphI)   GCAGC(8N)^(BbvI)
              GTT^AAC(HpaI)                                  CAGCAG(EcoP15)
              GTY^PAC(HindII)           ^GTNAC(MaeIII)     GC^NGC(Fnu4HI)
 ^(7N)TCTTC(MboII)                          CATG^(NlaIII)       GAAGA(8N)^(MboII)
                                                              ^(7N)GAGG(MnlI)

301       311       321       331       341       351       361       371       381       391
CAGGGCATGGGAGAGCTTTTTTGGACGGAATTCTTTCAGATGGATCTCCTAAATGCGAGTGCAATACTTGCTACACTGGTGCAGACTGCTCTGAAAAGATT
     CATG^(NlaIII)        G^AATTC(EcoRI)^GATC(MboI)
           AG^CT(AluI)   GAANN^NNTTC(XmnI)
                                GGATC(4N)^(BinI)
                               P^GATCY(XhoII)

401       411       421       431       441       451       461       471       481       491
ACAGGTTGCTCTGCTGATGTTGCCAGTGGTGATGGACTGTTTCTAGAAGAATACTGGCAGCAGCACAAGGAAAACAGTGCAGTGCTGGTTTCAGGATGG
                          GGTGA(8N)^(HphI)            GC^NGC(Fnu4HI)  AAC(6N)GTGC(EcoK)
                                    T^CTAGA(XbaI)     GCAGC(8N)^(BbvI)          (FokI)GGATG(9N)^
                                     C^TAG(MaeI)    CAGCAG(EcoP15)
                                        GAAGA(8N)^(MboII)

501       511       521       531       541       551       561       571       581       591
CACAGAATGAGCTACTTTTTCAACCCAGTTAGCAATTTCATATCTTTCGAGCTTGAAAAAACAATTAAAGAACTACATGAGATAGTCGGAAATGCTGCTG
        AG^CT(AluI)                           T^CGA(TaqI)                 CATG^(NlaII)
                                                                             (BbvI)^(12N)GCTGC
                                                                              (Fnu4HI)GC^NGC

601       611       621       631       641       651       661       671       681       691
CAAAGGACAGGTACATTGTGTTTGGAGTAGGGGTGACTCAACTCATCCATGGATTGGTCATCTCCCTTTCACCAAATATGACTGCCACTCCTTGTGCACC
          GT^AC(RsaI)          ^GTNAC(MaeIII)              ^(7N)TCACC(HphI)       (HgiAI)GWGCW^C
                                  G^ANTC(HinfI)                                    (SduI)GDGCHC
                               GGTGA(8N)^(HphI)
                                    ^(13N)CATCC(FokI)

701       711       721       731       741       751       761       771       781       791

ACAATCTAAAGTTGTTGCTCATGCCCCTTATTATCCGGTGTTTAGAGAACAAACAAAGTACTTTGACAAGAAAGGGTACGAGTGGAAAGGAAATGCAGCG
                  CATG^(NlaIII)  C^CGG(HpaII)   CAAPCA(Tth111-II)   (RsaI)GT^AC      (Fnu4HI)GC^NGC
                                                      AGT^ACT(ScaI)
                                                       GT^AC(RsaI)                      (BbvI)GCAGC(8N)^
                                                                                       (NspBII)CMG^CKG

801       811       821       831       841       851       861       871       881       891

GATTACGTGAACACTTCAACTCCAGAGCAATTCATTGAGATGGTTACTTCACCTAATAACCCAGAAGGTCTGCTTCGCCATGAAGTAATAAAGGGATGCA
    A^CGT(MaeII)  CTCCAG(GsuI)                                 GGTCT(EcoPI)              GGATG(9N)^(FokI)
                                                                                 CATG^(NlaIII)
                                                                                    (SfaNI)^(9N)GATG

901       911       921       931       941       951       961       971       981       991
AATCCATCTACGATATGGTTTACTACTGGCCTCATTACACCCCAATCAAGTACAAAGCCGATGAAGATATCATGTTGTTTACAATGTCTAAATACACTGG
                          WGG^CCW(HaeI)          GT^AC(RsaI)  GAAGA(8N)^(MboII)
                           GG^CC(HaeIII)                            CATG^(NlaIII)
                             CCTC(7N)^(MnlI)                    GAT^ATC(EcoRV)
```


1001 1011 1021 1031 1041 1051 1061 1071 1081 1091
ACACTCTGGTAGTCGATTTGGGTGGGCACTGATAAAGGATGAAACTGTGTATAATAAATTGTTGAATTACATGACAAAGAACACGGAGGGCACTTCCCGA
T^CGA(TaqI)
^(9N)TGGGTG(TaqII)
GDGCHC(SduI)
GGATG(9N)^(FokI)
CATG^(NlaIII)
(AvaI)^YCGPG
GDGCHC(SduI)
(MnlI)^(7N)GAGGC
1101 1111 1121 1131 1141 1151 1161 1171 1181 1191
GAAACACAGCTACGATCGCTCAAAATTCTAAAAGAAGTTATAGCAATGGTTAAAACACAGAAAGGCACCATGCGCGACCTCAACACATTTGGTTTTCAGA
AG^CT(AluI)
CGAT^CG(PvuI)
^GATC(MboI)
GGN^NCC(NlaIV)
G^GYPCC(HgiCI)
CATG^(NlaIII)
GCG^C(HhaI)
CG^CG(FnuDI)
CCTC(7N)^(MnlI)
1201 1211 1221 1231 1241 1251 1261 1271 1281 1291
AACTAAGAGAGAGGTGGGTAAATATCACTTCATTGCTCGATAAATCCGACAGATTCTCCTATCAAAAGCTTCCACAAAGTGAATATTGCAATTACTTCAG
C^TNAG(DdeI)
^(7N)GAGG(MnlI)
T^CGA(TaqI)
G^ANTC(HinfI)
A^AGCTT(HindIII)
AG^CT(AluI)
1301 1311 1321 1331 1341 1351 1361 1371 1381 1391
GAGAATGAGACCTCCATCCCCATCTTATGCATGGGTGAAGTGTGAATGGGAAGAAGACAAAGATTGCTACCAGACATTTCAGAATGGGCGTATCAATACG
AGACC(EcoPI)
CCTC(7N)^(MnlI)
^(13N)CATCC(FokI)
ATGCAT(AvaIII)
CATG^(NlaIII)
GGTGA(8N)^(HphI)
^(9N)TGGGTG(TaqII)
GAAGA(8N)^(MboII)
1401 1411 1421 1431 1441 1451 1461 1471 1481 1491
CAAAATGGAGAGGGTTTCGAAGCAGGTAGTCGTTATGTGCGTTTGAGTTTGATCAAGACAAAAGATGATTTTGATCAACTAATGTACTATTTGAAGAATA
^(7N)GAGG(MnlI)
TT^CGAA(AsuII)
T^CGA(TaqI)
^GATC(MboI)
T^GATCA(BclI)
^GATC(MboI)
(MboII)GAAGA(8N)^
GT^AC(RsaI)
1501 1511 1521 1531 1541 1551 1561 1571 1581 1591
TGGTTGAAGCAAAGAGGAAGACTCCTCTCATCAAACAACTTTCCAATGATCAGATTTCCCGCCGTCCTTTCATTTAAGTACTCATGTTATGTATTGCTCT
G^ANTC(HinfI)
^(7N)GAGG(MnlI)
CCTC(7N)^(MnlI)
CAAPCA(Tth111-II)
AGT^ACT(ScaI)
GT^AC(RsaI)
CATG^(NlaIII)
1601 1611 1621 1631 1641 1651 1661 1671 1681 1691
GCTGTTTTGTTAGTGTATGACTATGTTCATACATCCTAATGCTATGGTAGTAAGGAGTATCTTTCTATGCAATAAATAAAGTTCATGTTTGTGATCATGT
^(13N)CATCC(FokI)
(NlaIII)CATG^
T^GATCA(BclI)
^GATC(MboI)
CATG^(NlaIII)
1701 1711 1721 1731 1741 1751
ATGGGCTACTATGATTTTATAATAAAATCAATTTTCATATAAAAAAAAAAAAAAAAA

Pattern identifier	Pattern matched	Base number matched
AG^CT(AluI)	AGCT	126, 134, 230, 314, 510, 550 1108,1267.
TT^CGAA(AsuII)	TTCGAA	1416
C^YCGPG(AvaI)	CYCGPG	1096
ATGCAT(AvaIII)	ATGCAT	194, 1327
GCAGC(8N)^(BbvI)	GCAGC	261, 458, 461, 795
^(12N)GCTGC(BbvI)	GCTGC	25, 594, 597
T^GATCA(BclI)	TGATCA	150, 1472, 1547, 1692
GGATC(4N)^(BinI)	GGATC	342
CCA(5N)^NTGG(BstXI)	CCA(6N)TGG	149
C^TNAG(DdeI)	CTNAG	1203
AAC(6N)GTGC(EcoK)	AAC(6N)GTGC	474
CAGCAG(EcoP15)	CAGCAG	262, 459
AGACC(EcoPI)	AGACC	1308
GGTCT(EcoPI)	GGTCT	867
G^AATTC(EcoRI)	GAATTC	328
GAT^ATC(EcoRV)	GATATC	112, 966
GC^NGC(Fnu4HI)	GCNGC	25, 261, 458, 461, 594, 597, 795.
CG^CG(FnuDII)	CGCG	1173
GGATG(9N)^(FokI)	GGATG	495, 894, 1037
^(13N)CATCC(FokI)	CATCC	644, 1315, 1632
CTCCAG(GsuI)	CTCCAG	820
WGG^CCW(HaeI)	WGGCCW	927
GG^CC(HaeIII)	GGCC	928
GWGCW^C(HgiAI)	GWGCWC	694
G^GYPCC(HgiCI)	GGYPCC	1164
GCG^C(HhaI)	GCGC	1172
GTY^PAC(HindII)	GTYPAC	216
A^AGCTT(HindIII)	AAGCTT	1266
G^ANTC(HinfI)	GANTC	141, 635, 1252, 1520
GTT^AAC(HpaI)	GTTAAC	216
C^CGG(HpaII)	CCGG	735
GGTGA(8N)^(HphI)	GGTGA	242, 429, 632, 1334
^(7N)TCACC(HphI)	TCACC	669, 849
C^TAG(MaeI)	CTAG	444
A^CGT(MaeII)	ACGT	805
^GTNAC(MaeIII)	GTNAC	243, 633, 843
^GATC(MboI)	GATC	343, 1114, 1451, 1473, 1548,1693.
GAAGA(8N)^(MboII)	GAAGA	267, 447, 963, 1350, 1353, 1493 1517.
^(7N)TCTTC(MboII)	TCTTC	207
CCTC(7N)^(MnlI)	CCTC	29, 83, 930, 1178, 1311, 1524.
^(7N)GAGG(MnlI)	GAGG	270, 1086, 1211, 1410, 1514.
C^CATGG(NcoI)	CCATGG	647
CATG^(NlaIII)	CATG	47, 62, 203, 247, 306, 576, 648, 720, 879, 971, 1070, 1169, 1330, 1583, 1684, 1696.
GGN^NCC(NlaIV)	GGNNCC	1164
CMG^CKG(NspBII)	CMGCKG	796
CGAT^CG(PvuI)	CGATCG	1113
GT^AC(RsaI)	GTAC	611, 758, 776, 950, 1484, 1578.
AGT^ACT(ScaI)	AGTACT	757, 1577
GDGCHC(SduI)	GDGCHC	694, 1024, 1088
^(9N)GATGC(SfaNI)	GATGC	895.
T^CGA(TaqI)	TCGA	547, 1013, 1237, 1417

^(9N)TGGGTG(TaqII)	TGGGTG	1019, 1332
CAAPCA(Tth111-II)	CAAPCA	750, 1532
T^CTAGA(XbaI)	TCTAGA	443
P^GATCY(XhoII)	PGATCY	342
GAANN^NNTTC(XmnI)	GAA(4N)TTC	328

Patterns not matched.

GACGT^C(AmatII)	GACGTC	CGAAT(HinfIII)	CGAAT
GT^MKAC(AccI)	GTMKAC	ATTCG(HinfIII)	ATTCG
GP^CGYC(AcyI)	GPCGYC	GGTAC^C(KpnI)	GGTACC
C^TTAAG(AflII)	CTTAAG	A^CGCGT(MluI)	ACGCGT
A^CPYGT(AflIII)	ACPYGT	TGC^GCA(MstI)	TGCGCA
TTT^AAA(AhaIII)	TTTAAA	GCC^GGC(NaeI)	GCCGGC
GGGCC^C(ApaI)	GGGCCC	GG^CGCC(NarI)	GGCGCC
G^GNCC(AsuI)	GGNCC	CA^TATG(NdeI)	CATATG
G^GWCC(AvaII)	GGWCC	GC^GGCCGC(NotI)	GCGGCCGC
C^CTAGG(AvrII)	CCTAGG	TCG^CGA(NruI)	TCGCGA
TGG^CCA(BalI)	TGGCCA	PCATG^Y(Nsp[7524]I)	PCATGY
G^GATCC(BamHI)	GGATCC	CTGCA^G(PstI)	CTGCAG
GCC(4N)^NGGC(BglI)	GCC(5N)GGC	CAG^CTG(PvuII)	CAGCTG
A^GATCT(BglII)	AGATCT	GAGCT^C(SacI)	GAGCTC
^(5N)GATCC(BinI)	GATCC	CCGC^GG(SacII)	CCGCGG
GCGCGC(BsePI)	GCGCGC	G^TCGAC(SalI)	GTCGAC
G^GTNACC(BstEII)	GGTNACC	CC^TNAGG(SauI)	CCTNAGG
CC^SGG(CauII)	CCSGG	CCNGG(ScrFI)	CCNGG
PCCGGY(Cfr10I)	PCCGGY	GCATC(5N)^(SfaNI)	GCATC
Y^GGCCP(CfrI)	YGGCCP	GGCC(4N)^NGGCC(SfiI)	GGCC(5N)GGCC
AT^CGAT(ClaI)	ATCGAT	CCC^GGG(SmaI)	CCCGGG
TTT^AAA(DraI)	TTTAAA	TAC^GTA(SnaBI)	TACGTA
AGCGCT(Eco47III)	AGCGCT	GTATAC(SnaI)	GTATAC
GAG(7N)GTCA(EcoA)	GAG(7N)GTCA	GCATG^C(SphI)	GCATGC
TGA(8N)TGCT(EcoB)	TGA(8N)TGCT	AGG^CCT(StuI)	AGGCCT
ATCA(7N)ATTC(EcoDXI)	ATCA(7N)ATTC	GACCGA(11N)^(TaqII)	GACCGA
^CCWGG(EcoRII)	CCWGG	^(9N)TCGGTC(TaqII)	TCGGTC
Y^GGCCG(GdiII)	YGGCCG	CACCCA(11N)^(TaqII)	CACCCA
PGCGC^Y(HaeII)	PGCGCY	GACN^NNGTC(Tth111-I)	GAC(3N)GTC
GACGC(5N)^(HgaI)	GACGC	C^TCGAG(XhoI)	CTCGAG
^(10N)GCGTC(HgaI)	GCGTC	C^CCGGG(XmaI)	CCCGGG
ACC(6N)GGT(HgiEII)	ACC(6N)GGT	C^GGCCG(XmaIII)	CGGCCG
GPGCY^C(HgiJII)	GPGCYC		

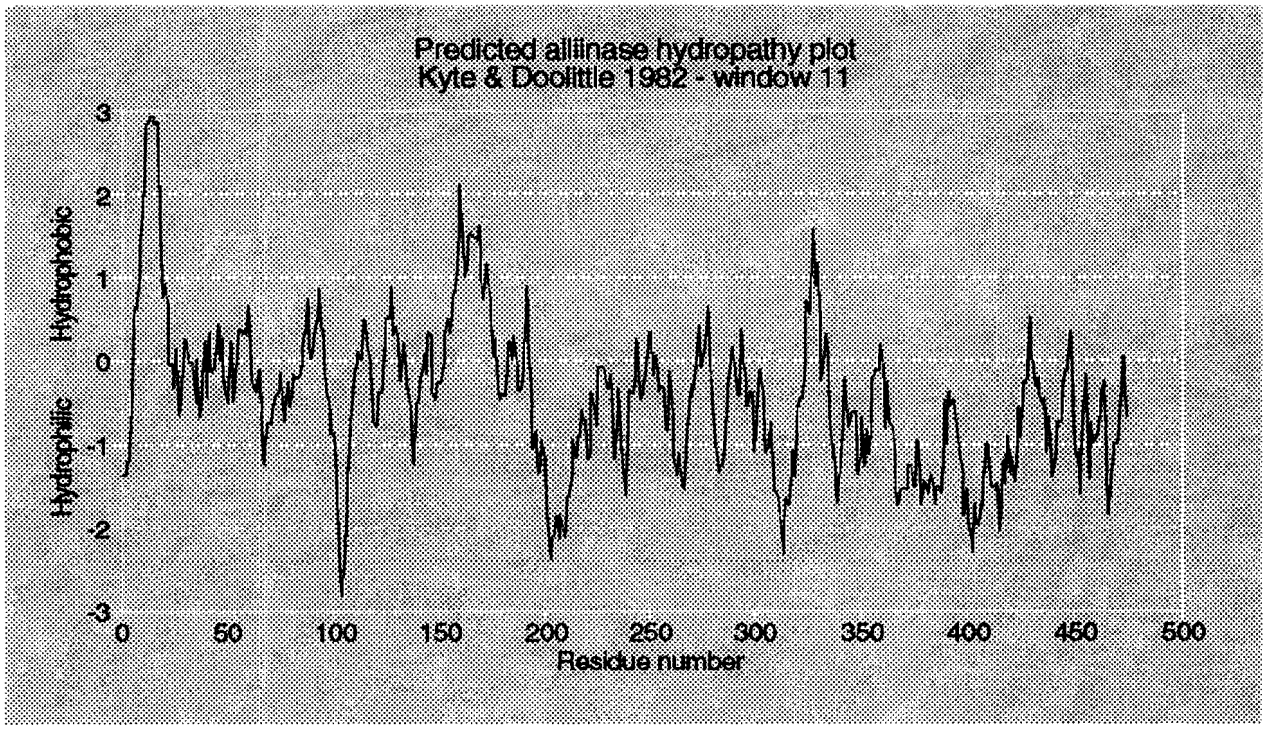
N = A or T or C or G, Y = pyrimidine, P = purine, W = A or T, M = A or C, S = G or C, K = G or T, H = A or C or T, D = G or A or T.

Figure 3.10. Hydrophobicity and antigenicity plots constructed from the amino acid sequence predicted from the *A. cepa* alliinase cDNA

Panel A. The programs SeqAid and Genesys were used to analyse the hydrophobic/hydrophilic nature of the predicted amino acid sequence of *A. cepa* alliinase. Hydrophobicity analysis is according to the protocol of Kyte and Doolittle (1982). Values averaged over a window of 11 residues are plotted at the sixth residue.

Panel B. Predicted antigenicity of the deduced alliinase amino acid sequence was analysed according to the method of Welling *et al.* (1985). Values averaged over a window of 7 residues are plotted at the position of the fourth residue.

A.


Predicted alliinase hydropathy plot
Kyte & Doolittle 1982 - window 11
Hydrophobic
Hydrophilic
3
2
1
0
-1
-2
-3
0
50
100
150
200
250
300
350
400
450
500
Residue number

B.

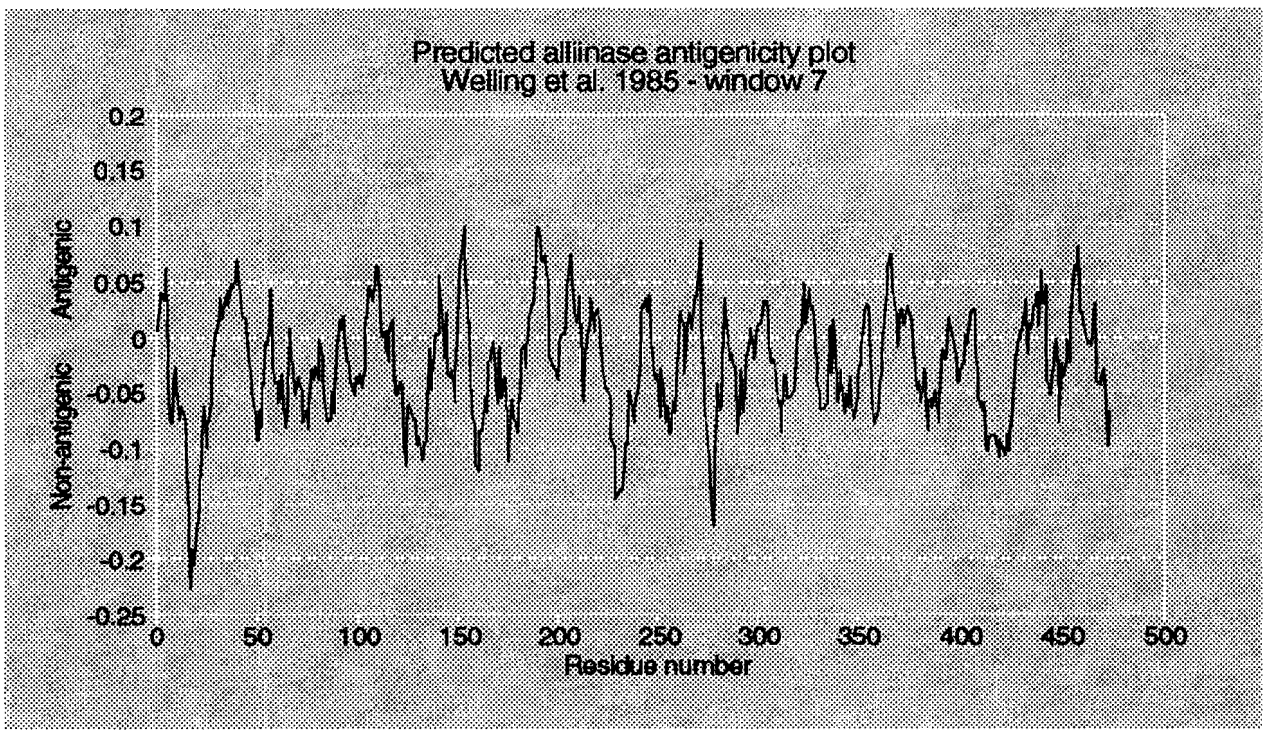

Predicted alliinase antigenicity plot
Welling et al. 1985 - window 7
Antigenic
Non-antigenic
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
0
50
100
150
200
250
300
350
400
450
500
Residue number

Figure 3.11. Secondary structure prediction according to the protocol of Garnier *et al.* (1978) of the amino acid sequence inferred by translation of the *A. cepa* alliinase cDNA

The PlotStructure and PeptideStructure programs of the University of Wisconsin Genetics Computer Group Sequence Analysis Package were used to generate this predicted structure. Small sawtooth, sine waves and soft undulating waves (barely visible at this scale) are used to denote β-strands, α-helices and random coils respectively. O refers to Asn-linked glycosylation sites, Cys refers to cysteine residues.

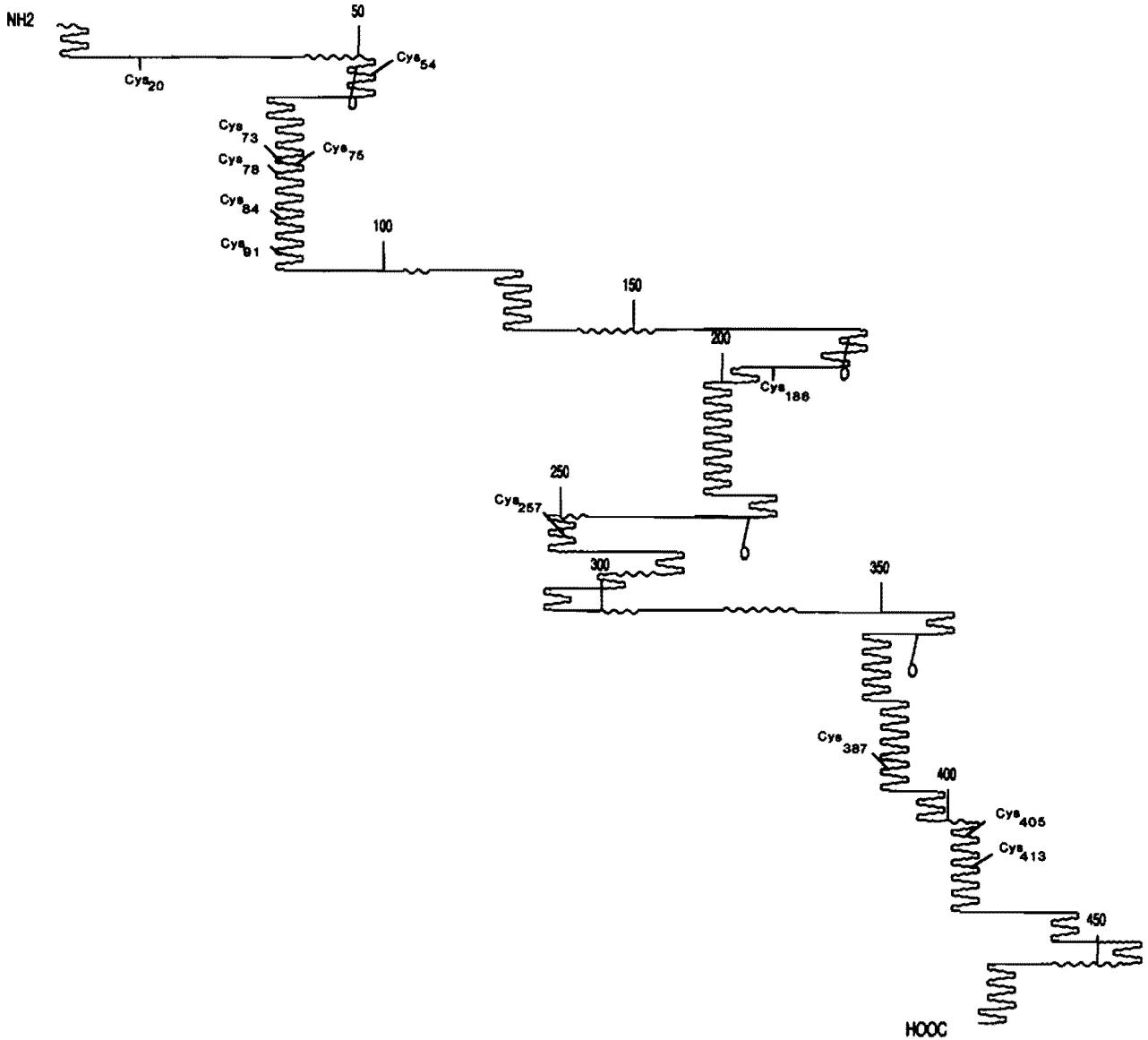

NH2
50
Cys20
Cys54
Cys73
Cys75
Cys78
Cys84
Cys91
100
150
200
Cys186
250
Cys257
300
350
Cys387
400
Cys405
Cys413
450
HOOC

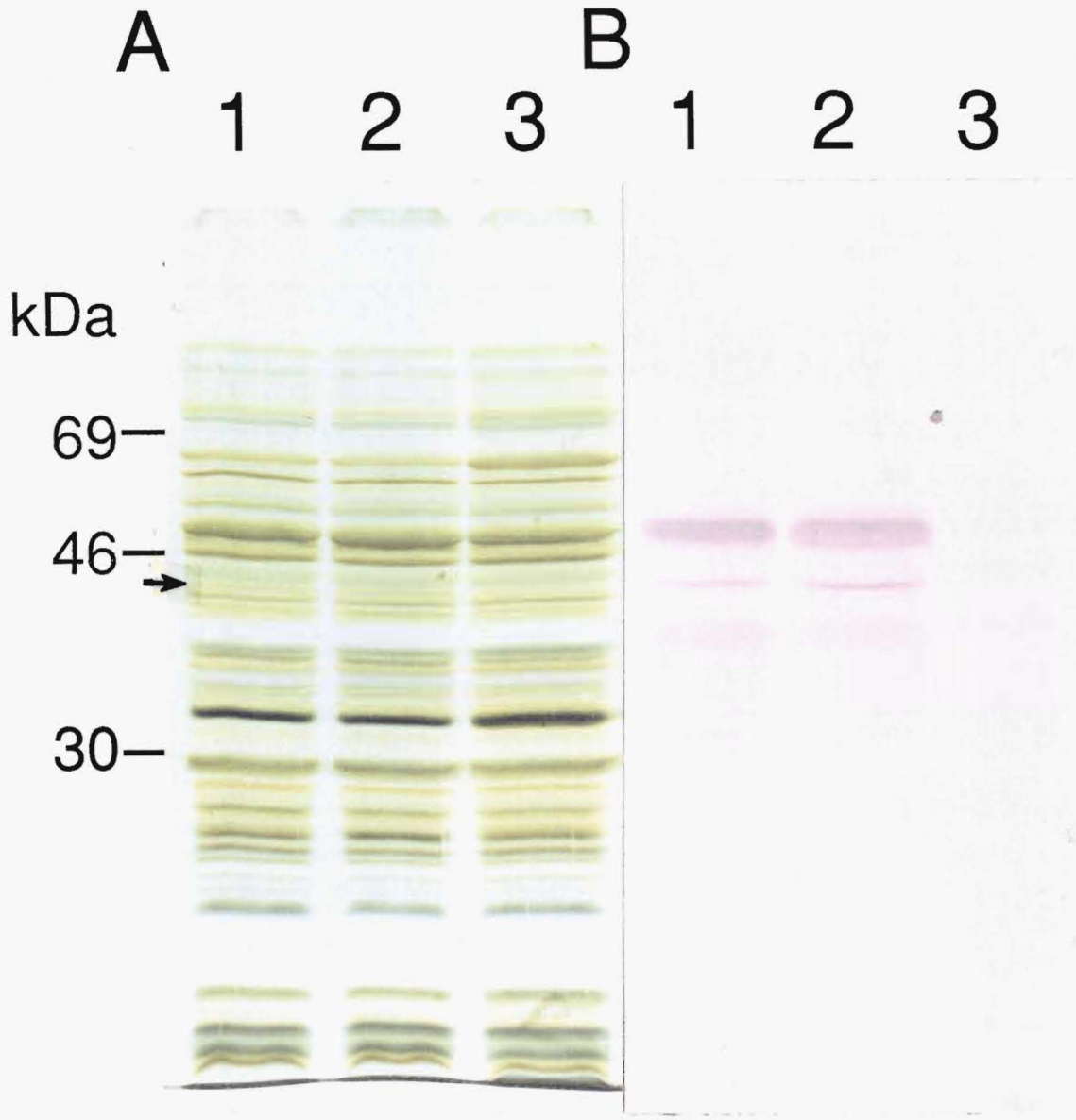
Figure 3.12. Western blot analysis of the protein expressed by the alliinase clone Alli6 in *E. coli*

Total protein from *E. coli* XL1-Blue containing the pAlli6 phagemid (XLalli6) or pBluescript(-) (XL(-)) were separated on denaturing 12% polyacrylamide gels (see Subsection 2.4.6.6). Duplicate gels were either **A.** silver stained, or **B.** blotted onto nitrocellulose membranes (Hybond-C, Amersham). The membrane was probed with I66 anti-alliinase antisera at 1:8 000 dilution and developed as described in Subsection 2.4.2. The arrow indicates the position of the 41.4 kDa immunoreactive protein visible among the protein profile of XLalli6.

Lane 1. XLalli6 grown in the presence of 1 mM IPTG.

Lane 2. XLalli6 grown without IPTG exposure.

Lane 3. XL(-) grown in the presence of 1 mM IPTG.


A
B
1
2
3
1
2
3
kDa
69
46
30

3.7. Characterization of the genomic organization of alliinase genes

Southern analysis was performed to ascertain the number of genes encoding *A. cepa* (cultivar Pukekohe Longkeeper) alliinase. *A. cepa* bulb DNA was digested with a range of restriction endonucleases with 6 bp recognition sites. Digestion with the enzymes *Cla* I, *Kpn* I, *Nru* I, *Pst* I, *Sal* I and *Xho* I resulted in little fragmentation of *A. cepa* DNA after overnight incubation, although examination on agarose gels of *A. cepa* DNA digested with the restriction enzymes *Bam* HI, *Bgl* II, *Dra* I, *Eco* RI, *Eco* RV, *Hin*d III and *Xba* I showed considerable digestion had occurred (data not presented).

To establish whether the limited extent of fragmentation with *Cla* I, *Kpn* I, *Nru* I, *Pst* I, *Sal* I and *Xho* I may be accounted for by the presence of a contaminating endonuclease inhibitor, samples of *A. cepa* DNA were further purified. This involved two phenol extraction steps, two chloroform extraction steps and a further ethanol precipitation. The endonuclease digestions were then repeated with twice the ratio of enzyme to DNA used previously, and incubated overnight. Further enzyme was added and the samples were again re-incubated. However neither these attempts, nor the use of independently isolated *A. cepa* DNA samples improved the extent of cleavage. Southern analysis (described in Section 2.6) of these samples resulted in an absence of discrete signals and a smear of hybridization at the top of each lane, similar to results obtained with undigested DNA. Consequently the restriction enzymes *Bam* HI, *Bgl* II, *Dra* I, *Eco* RI, *Eco* RV, *Hin*d III and *Xba* I were used for all subsequent Southern analysis. No sites for *Bam* HI, *Bgl* II or *Dra* I are present within the sequence of the alliinase cDNAs isolated. These experiments were duplicated with DNA extracted from two *A. cepa* bulbs of the cultivar Pukekohe Longkeeper (bulb A and bulb B).

The full-length cDNA probe, Alli6, hybridized to multiple *A. cepa* DNA fragments under conditions of high stringency (65°C, 0.1x SSC wash) as shown in Figure 3.13 & 3.14 (pages 137 and 139). Hybridization with *Bam* HI-digested DNA, labelled three fragments (one of 10.7 kb, and others of 8.8 kb and 5.7 kb). A dilution of Alli6 probe DNA representing 1.1 x 10^{-5}% of the genomic DNA present in each lane, was loaded in a separate lane of each gel to approximate the relative abundance of a 1.7 kbp single copy gene in the *A. cepa* genome (described in Subsection 2.6.4). The single copy reconstruction present on the membrane used to generate the autoradiograph shown in Figure 3.13 was obscured by a background hybridization signal, however comparison of band intensities to single copy

reconstructions were made from other replicate blots not shown here. Comparison can also be made to the 4.0 kb fragment present in the *Dra* I digest, whose intensity was determined from replicate Southern blots to approximate that of a single copy reconstruction. These comparisons suggest that the two largest *Bam* HI bands (10.7 kb and 8.8 kb) hybridized to the Alli6 probe with intensities greater than that of a single copy gene, whereas the smaller band was of lower intensity than a single copy. These results imply that more than three homologous gene copies are present within the *A. cepa* genome.

Bgl II-digestion of *A. cepa* DNA contained four hybridizing bands (7.1 kb, 5.4 kb, 4.4 kb and 3.1 kb) with intensities approximately equivalent to a single copy, another that was slightly fainter at 8.6 kb, and also one band of 4.0 kb that was very faint (Figure 3.13). Hybridization with *Xba* I-digested *A. cepa* DNA (Figure 3.14) labelled four bands (10 kb, 6.6 kb, 5.6 kb and 3.0 kb). The three smaller ones showed hybridization intensities equal to or greater than single copy, while the larger band was fainter. These results also suggest more than three homologous gene copies are present.

Hybridization analysis of *Dra* I-digested *A. cepa* DNA samples extracted from two bulbs resulted in slightly different hybridization patterns. DNA extracted from one onion (bulb B, Figure 3.14, Panel A) produced eight strongly hybridizing bands (4.7 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.1 kb, 1.3 kb, 1.2 kb and 1.1 kb) and a single faint band (4.0 kb) when hybridized with Alli6. Digestion with *Dra* I of the DNA extracted from bulb A (Figure 3.13) resulted in a hybridization pattern which lacked the 4.7 kb band but instead displayed a strong 4.0 kb band and two other faint bands (1.0 kb and 0.65 kb). The binding of the 4.7 kb and 4.0 kb bands were of equivalent intensity to single copy reconstructions but the other bands appeared to be slightly weaker. The incomplete resolution of the bands between 2.7 kb and 2.9 kb made accurate determination of their hybridization intensities difficult. *Eco* RI digestion of bulb B DNA (Figure 3.14, Panel A) produced five bands of about single copy intensity (7.0 kb, 4.8 kb, 4.0 kb, 3.7 kb and 3.1 kb). Digestion of DNA from bulb A with *Eco* RI (Figure 3.13) gave a clearly different hybridization pattern, with five bands of intensity about that of single copy (7.2 kb, 4.1 kb, 3.9 kb, 3.3 kb, and 3.0 kb) and three fainter bands of 4.6 kb, 2.6 kb and 1.74 kb.

Hybridization of the Alli6 probe to *Hin*d III-digested DNA produced three bands of about single copy intensity (4.9 kb, 1.6 kb and 1.05 kb) and seven fainter bands (4.4 kb, 4.0 kb, 3.0 kb, 2.5 kb, 2.2 kb, 2.0 kb and 1.7 kb) indicating more than three homologous genomic equivalents are present. *Eco* RV digestion (Figure 3.14)

produced one band of about single copy intensity (3.2 kb), eight bands (7.6 kb, 6.3 kb, 5.7 kb, 5.0 kb, 2.9 kb, 2.6 kb, 2.3 kb and 2.0 kb) that hybridized at slightly lower intensity, and one band (4.3 kb) that was notably weaker. Pea DNA digested with *Hind* III was included on Southern blots probed with Alli6 to ensure hybridization was specific to *A. cepa* genomic sequences. Hybridization to two fragments of 6 kb and 5 kb was faintly discernable (data not shown).

To establish the relatedness of the multiple genomic fragments complementary to the Alli6 cDNA, and whether there is more than one gene transcriptionally active in sprouting leaf tissue of *A. cepa* bulbs, Southern blots were stripped of previously bound probe (see Subsection 2.6.3) and re-hybridized with a fragment containing the 3' untranslated region of Alli6 (Alli6-3') (described in Subsection 2.6.2). Stripping the membranes with sodium hydroxide lowered the sensitivity of the hybridization, possibly due to removal of some of the *A. cepa* DNA. To compensate for this membranes were exposed to films for two weeks, which resulted in a higher grey background. This 3' probe hybridized most strongly to a single band in each lane. These were the 3.2 kb *Eco* RV fragment, the 4.0 kb *Eco* RI fragment of bulb A, the 6.6 kb *Xba* I fragment and the 2.9 kb *Dra* I fragment (Figure 3.14, Panel B), as well as the 2 kb *Hind* III fragment and the 5.4 kb *Bgl* II fragment (data not presented). Other bands were also detected at lower intensity, particularly the 5.0 kb and 2.0 kb *Eco* RV bands, the 7.0 kb and 3.1 kb *Eco* RI bands, the 6.6 kb *Xba* I band and the 1.3 kb *Dra* I band. Some of the bands detected by the Alli6 probe were not detected by the Alli6-3' probe, notably the three largest *Eco* RV bands, the 3.7 kb *Eco* RI band, the 10 kb *Xba* I band and the 4.7 kb, 2.1 kb, 1.2 kb and 1.1 kb *Dra* I bands.

To gain an indication of whether the *A. cepa* genome contained further alliinase sequences with less sequence similarity to the Alli6 cDNA probe, Southern blots were probed with Alli6 under identical conditions as before, but were washed only at low stringency (2x SSC, 65°C). Under these conditions, the probe would be expected to hybridize to sequences of comparatively lower homology. Figure 3.15 shows the autoradiograph produced using DNA extracted from bulb A. No differences were found in the number of bands produced by *Bgl* II (Figure 3.15), *Bam* HI or *Xba* I (not shown). An extra faint band at 1.6 kb was detected within *Dra* I-digested *A. cepa* DNA (Figure 3.15). *Eco* RI-digested *A. cepa* DNA produced several more faint bands of 4.3 kb, 3.3 kb, 2.6 kb and 1.7 kb. There were also four other very faint bands which were just visible in the *Eco* RI lane, of 5.4 kb, 2.2 kb, 2.1 kb and 1.9 kb.

Figure 3.13. Estimation of the *A. cepa* alliinase gene family size

A. cepa DNA was extracted from bulb tissue and 60 μg digested with a range of restriction enzymes. Digested DNA was separated on 1.2% agarose gels in 1x TBE (described in Subsection 2.6.3) and blotted onto nylon membranes (Zetaprobe, Biorad). Membranes were hybridized with ^{32}P-labelled Alli6 DNA (see Subsection 2.6.2) overnight at 65°C. Wash conditions were 30 minutes at 65°C each in 2x SSC + 0.1% SDS, 1x SSC + 0.1% SDS, and 0.1x SSC + 0.1% SDS. Autoradiographs were produced by exposure of membranes to Kodak X-omat AR film with intensifying screens for 1 week at -70°C.

Presented are the results from bulb A DNA digested with;

DI - *Dra* I, **EI** - *Eco* RI, **HIII** - *Hind* III, **BII** - *Bgl* II, **BH** - *Bam* HI.

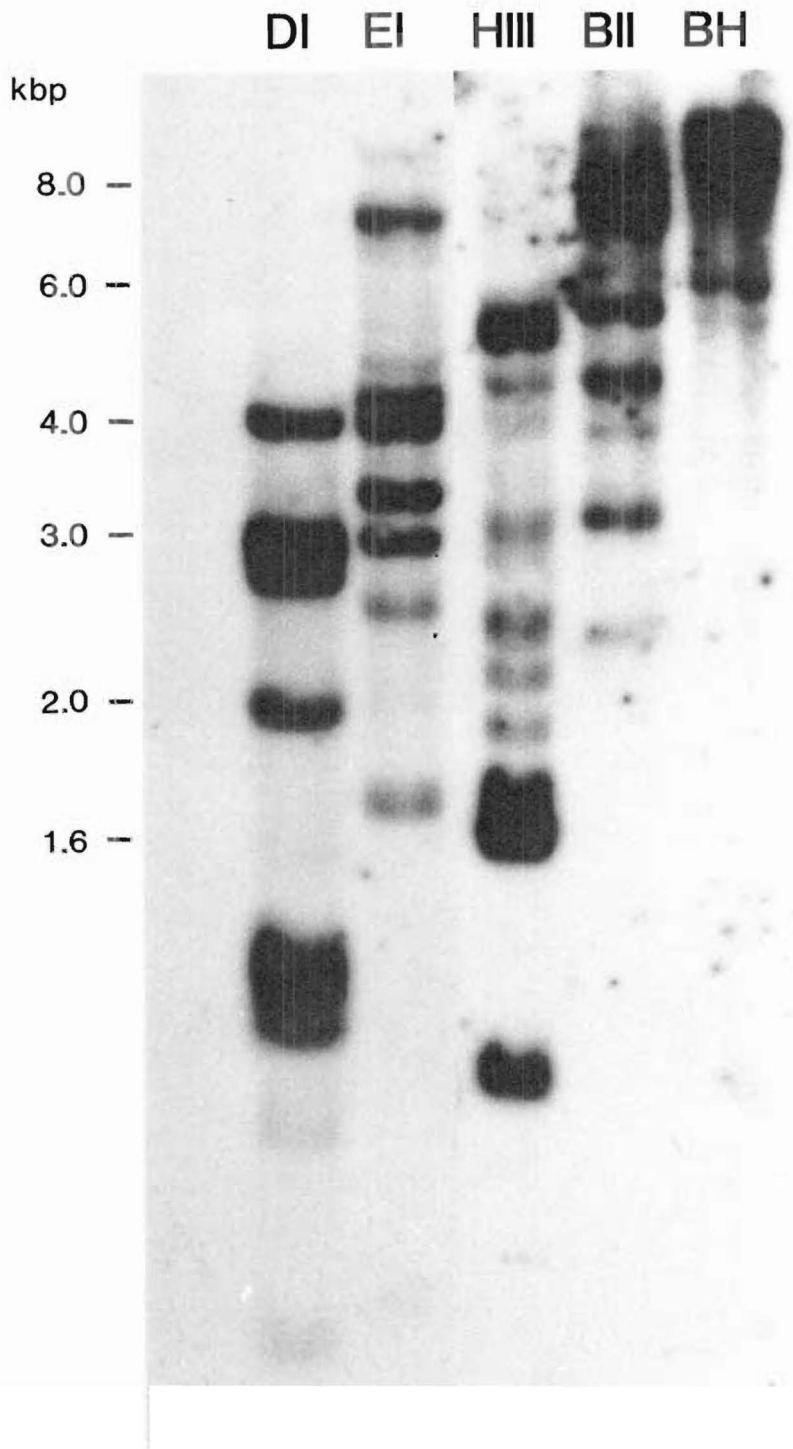

DI
EI
HIII
BII
BH
kbp
8.0
6.0
4.0
3.0
2.0
1.6

Figure 3.14. Identification of the alliinase gene transcriptionally active in *A. cepa* bulb sprouting leaf tissue

A. cepa DNA was extracted from bulb tissue and 60 μg digested with a range of restriction enzymes. Digested DNA was separated on 1.2% agarose gels in 1x TBE (described in Subsection 2.6.3) and blotted onto nylon membranes (Zetaprobe, Biorad). Membranes were hybridized with random primed ^{32}P-labelled Alli6 DNA (see Subsection 2.6.2) overnight at 65°C. Wash conditions were 30 minutes each in 2x SSC, 1x SSC, and 0.1x SSC at 65°C. Autoradiographs were produced by exposure of membranes to Kodak X-omat AR film with intensifying screens for 1 week (Panel A) or 2 weeks (Panel B) at -70°C.

Presented are the results from bulb B DNA digested with;

EV - *Eco* RV, **EI** - *Eco* RI, **XI** - *Xba* I, **DI** - *Dra* I.

1 - Alli6 probe DNA (6 pg) was loaded to represent a single copy gene (described in Subsection 2.6.4).

Panel A. The membrane was probed with Alli6.

Panel B. The membrane was stripped and reprobed with Alli6-3'.

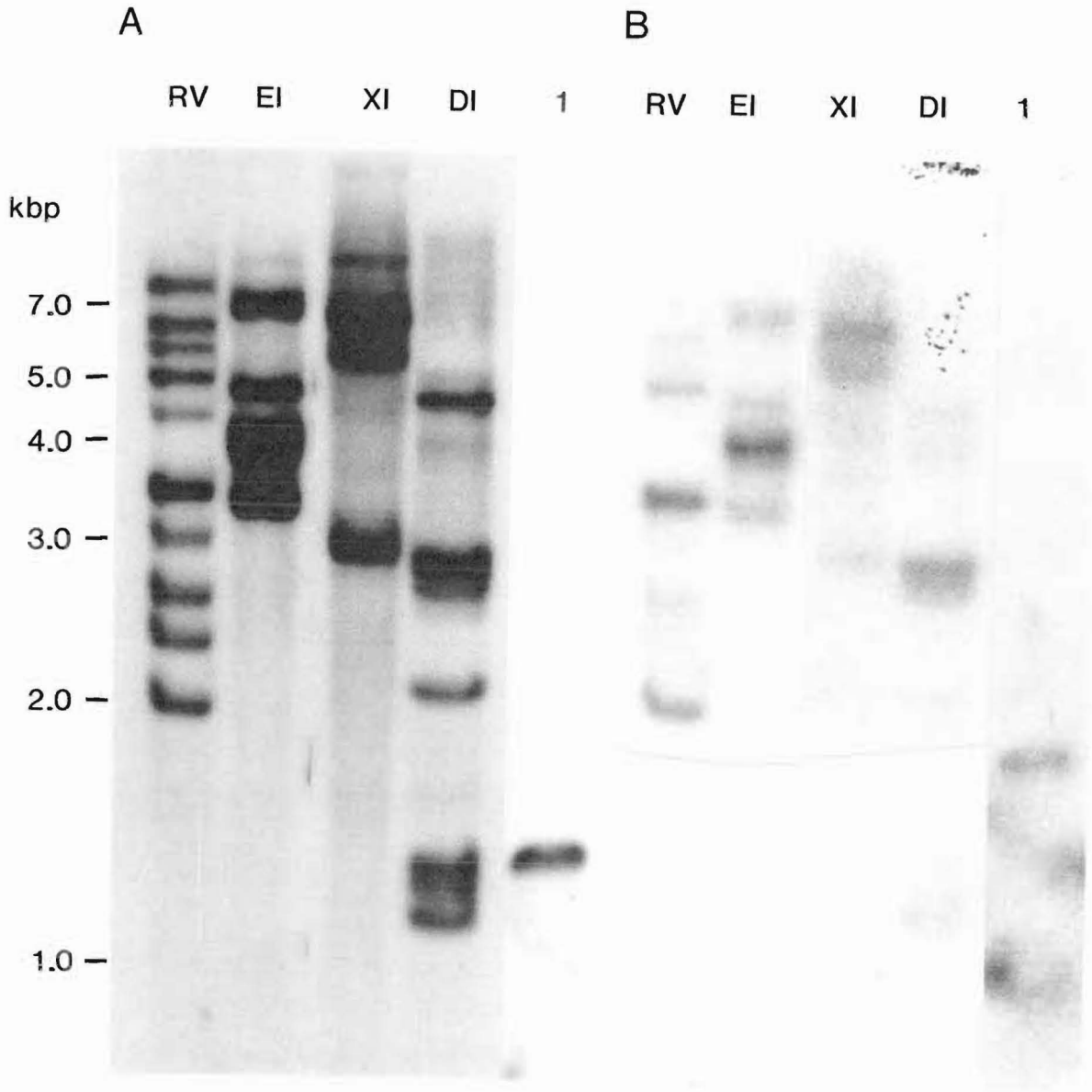

A
RV
EI
XI
DI
1
kbp
7.0
5.0
4.0
3.0
2.0
1.0
B
RV
EI
XI
DI
1

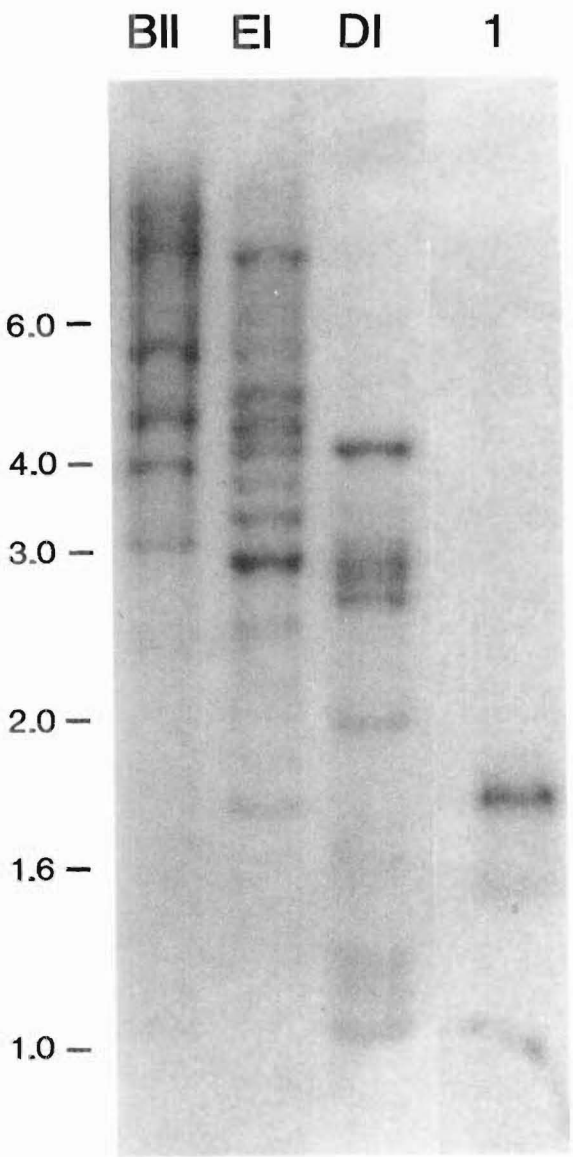
Figure 3.15. Estimation of the size of the *A. cepa* alliinase gene family

A. cepa DNA was extracted from bulb tissue and 20 μg digested with a range of restriction enzymes. Digested DNA was separated on 1.2% agarose gels in 1x TBE (described in Subsection 2.6.3) and blotted onto nylon membranes (Zetaprobe, Biorad). Membranes were hybridized with Alli6 labelled by random priming (see Subsection 2.6.2) overnight at 65°C. Wash conditions were 30 minutes in 2x SSC. Autoradiographs were produced by exposure of membranes to Kodak X-omat AR film with intensifying screens for 1 week at -70°C.

Presented are the results from bulb B DNA digested with;

BII - *Bgl* II, **EI** - *Eco* RI, **DI** - *Dra* I.

1 - single copy reconstruction, 2 pg of Alli6 probe DNA.


BII
EI
DI
1
6.0
4.0
3.0
2.0
1.6
1.0

3.8. Expression of alliinase genes during *A. cepa* seedling development

To determine if the expression of alliinase might be regulated at the level of transcription, northern analysis was performed using RNA from developing *A. cepa* seedlings. This tissue was selected for study as variation in alliinase activity within developing *A. cepa* seedlings had been reported (Freeman 1979). RNA was extracted from *A. cepa* seedlings grown under light and dark regimes and separated into five stages of development up to and including the time of emergence of the first true leaf (see Subsection 2.7.1). Ungerminated seeds were not included in the analysis as trial extractions did not yield RNA of sufficient yield or quality.

RNA was quantitated spectrophotometrically by measuring absorbance at a wavelength of 260 nm (Sambrook *et al.* 1989). Subsequent visual examination of theoretically equivalent loadings separated by formaldehyde-agarose gel electrophoresis suggested that the spectrophotometric estimations of RNA concentration may be inaccurate (Figure 3.16, Panel A, page 146). To correct for this, and for potential variation in the efficiency of RNA transfer to membranes, all northern blots were probed with a clone containing genes for the *Malus* 4.5S, 18S and 26S rRNAs. The intensity of hybridization of this 45S rDNA to each lane was quantitated by scanning each lane of the resulting autoradiographs with a Shimadzu densitometer (see Subsection 2.7.5). These integrated optical density (IOD) values were then used to determine the degree of variation in the quantity of RNA present from each sample.

The Alli6 cDNA was used as a representative probe to ascertain the total abundance of alliinase-encoding transcripts in *A. cepa* seedling RNA. Hybridizations were carried out under high stringency (65°C, with final washes at 0.1x SSC, 0.1% SDS). Under these conditions the Alli6 probe hybridized to a single band in each tissue sample, indicating that alliinase transcripts were present in *A. cepa* seedlings at all five stages of development tested and in both light- and dark-grown tissues. By comparison with RNA molecular size standards, this band was calculated to represent a transcript of 1.7 kb (Figure 3.16, Panel B). This transcript was easily detected in 10 μg of *A. cepa* RNA, with strong autoradiographic signals being produced on X-ray film after only 4 to 5 hours exposure (Figure 3.16, Panel B). No binding of the probe to total potato RNA was detected, confirming that hybridization was specific to an *A. cepa* transcript.

Variation in the hybridization intensities suggested a change in transcript abundance occurred as seedlings developed. To quantify this variation, IOD values of signals were determined from autoradiographs produced from four separate northern blot membranes probed with Alli6. Autoradiographs resulting from 45S rDNA probing of these same membranes were also scanned using identical parameters. To determine if the differences in the binding intensities were related to possible RNA loading or transfer variation, correlation coefficients between the IOD values obtained for the Alli6 probe and those of the 45S rDNA were calculated. These data are presented in Table 3.7 (page 144). The calculated correlation coefficients(r^2) were less than 0.16 for three of the four northerns analysed, indicating the binding intensity of the Alli6 probe was not solely related to the total amount of RNA present on the membranes. The fourth northern blot analysed in this way showed a correlation of 0.4, suggesting RNA loadings may be influencing the pattern found.

To correct for variation in the alliinase IOD values that may be attributable to loading or transfer variation, IOD values of Alli6-generated signals were corrected relative to those determined after hybridization of the same membranes with the rDNA probe. As shown in Table 3.7, this correction significantly lowered the correlation coefficients, with no values above 0.16. This indicates the pattern of alliinase transcript abundance found was not solely attributable to RNA loading and transfer variation. The overall result from the four northerns analysed is shown graphically in Figure 3.17 (page 146), in which the data are presented so that the lowest corrected IOD value is made equal to 1 and the others are calculated proportionally.

This analysis suggests that alliinase transcript abundance is high soon after germination. Alliinase mRNA levels appear to peak at Stage 2, approximately 2 to 3 days after imbibition (DAI), which correlates with the emergence of the *A. cepa* root tip and hypocotyl. The relative abundance of these transcripts was also high at Stage 3, as the cotyledon developed, but dropped rapidly during cotyledon elongation (Stage 4) and was significantly lower by the time of emergence of the first true leaf (Stage 5). This pattern was almost identical in the light- and dark-grown tissues, although the decline in the abundance of alliinase transcripts at stage 4 was more marked in dark-grown seedlings. In general, dark-grown seedlings showed a slightly lower relative abundance of alliinase mRNA than light-grown tissues, and this difference was found to be statistically significant by analysis of variance ($p < 0.05$).

Table 3.7. Integrated optical density (IOD) values obtained from northern analysis of *A. cepa* seedling RNA probed with Alli6 cDNA and *Malus* 45S rDNA

Integrated Optical Density (IOD) values												
	Northern blot #1			Northern blot #2			Northern blot #3			Northern blot #4		
Sample	Alli6		45S rDNA	Alli6		45S rDNA	Alli6		45S rDNA	Alli6		45S rDNA
	Measured	Corrected		Measured	Corrected		Measured	Corrected		Measured	Corrected	
[1]L1	175368	127970	245671	113526	103289	251181	555057	703335	338489	38274	37941	455822
L2	191729	152759	225005	166882	159723	238775	468492	499046	402653	78549	81337	436372
L3	202564	161731	224534	135652	134620	230283	574752	560855	439541	111745	78459	643561
L4	101073	175109	103476	74537	82426	206658	452797	507853	382415	15935	27211	264610
L5	32130	45627	126241	7923	14956	121067	171863	190153	387657	5342	6069	397732
D1	142920	210343	121809	43179	61581	160241	32600	355624	449387	39391	35704	498526
D2	180624	172328	187903	112731	109296	235715	563801	524034	461462	55546	52670	476533
D3	153774	195536	140984	89677	115431	177544	562744	412561	585048	34302	41779	370997
D4	56566	46421	218452	69390	35123	451494	58436	74972	334310	18036	15658	520471
D5	62925	56787	198651	12023	12939	212358	273811	231106	508170	11041	10989	453987
[§]r^2	0.15	0.08		0.09	0.004		0.11	0.004		0.40	0.15	

[1] Samples with the prefix 'L' are light-grown, those with the prefix 'D' are dark-grown.
Stage 1 = 1-2 days after imbibition (DAI); Stage 2 = 2-3 DAI; Stage 3 = 3-6 DAI; Stage 4 = 6-9 DAI; Stage 5 = 10-14 DAI.

[§] Linear correlation coefficients calculated between the IOD values for Alli6 and rDNA probing.

Figure 3.16. Expression analysis of alliinase transcription in developing *A. cepa* seedlings

Total RNA was extracted from *A. cepa* seedlings at five stages of development (see Subsection 2.7.1) and from *S. tuberosum* meristem tissue. RNA (10 μg) was fractionated under denaturing conditions (see Subsection 2.4.4.1.1) and blotted onto Zetaprobe membranes (Biorad). Membranes were hybridized overnight with Alli6 probe labelled by random priming, then washed and visualized by autoradiography (see Subsection 2.7.3).

Panel A. Ethidium bromide-stained gel prior to blotting.

Panel B. Autoradiograph of northern-blotted gel probed with Alli6.

Samples labelled 'L' are light-grown, those labelled 'D' are dark-grown.

Stage 1 = 1-2 days after imbibition (DAI)
Stage 2 = 2-3 DAI
Stage 3 = 3-6 DAI
Stage 4 = 6-9 DAI
Stage 5 = 10-14 DAI
P = *S. tuberosum* RNA
M = RNA molecular size markers

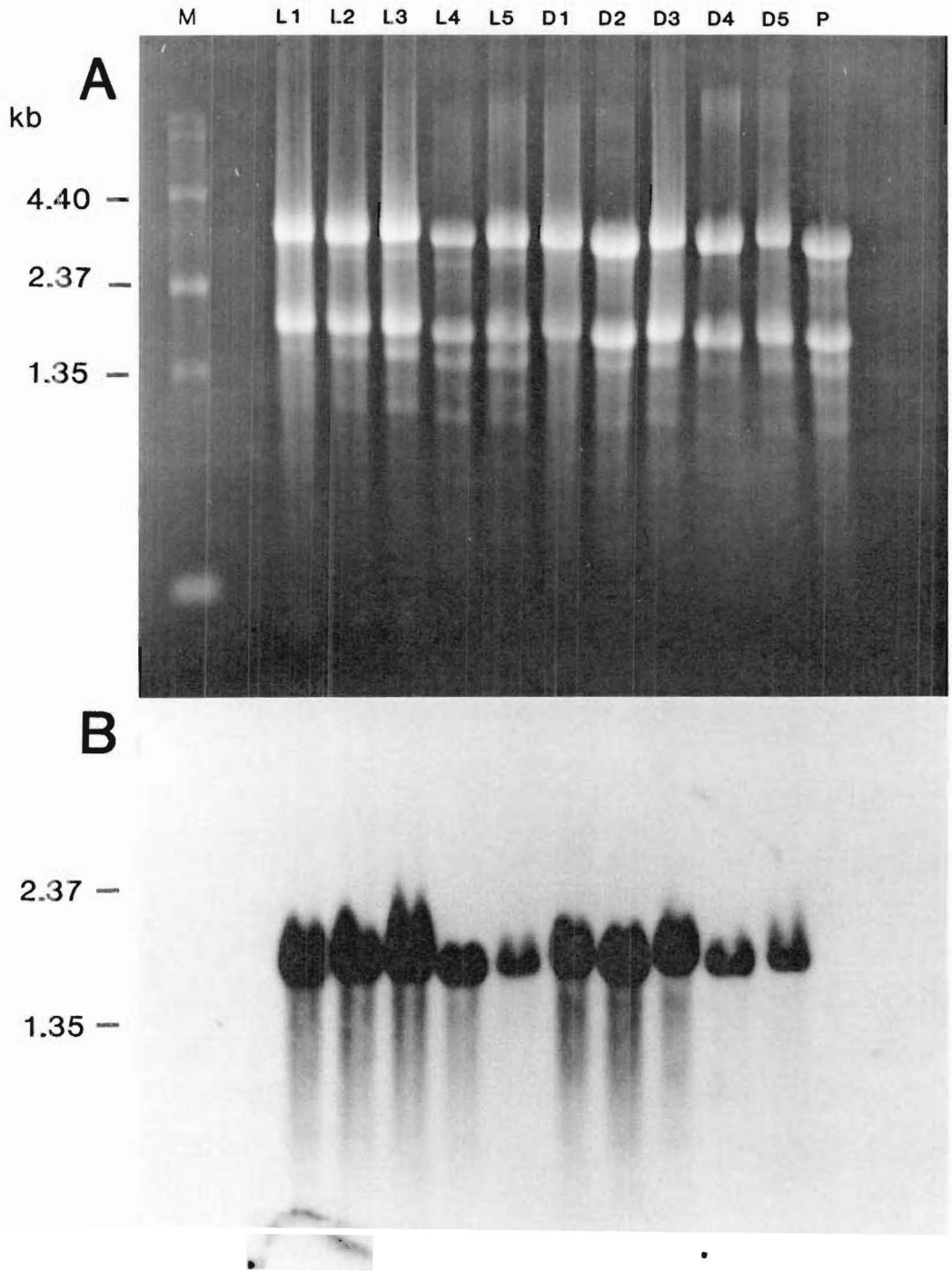
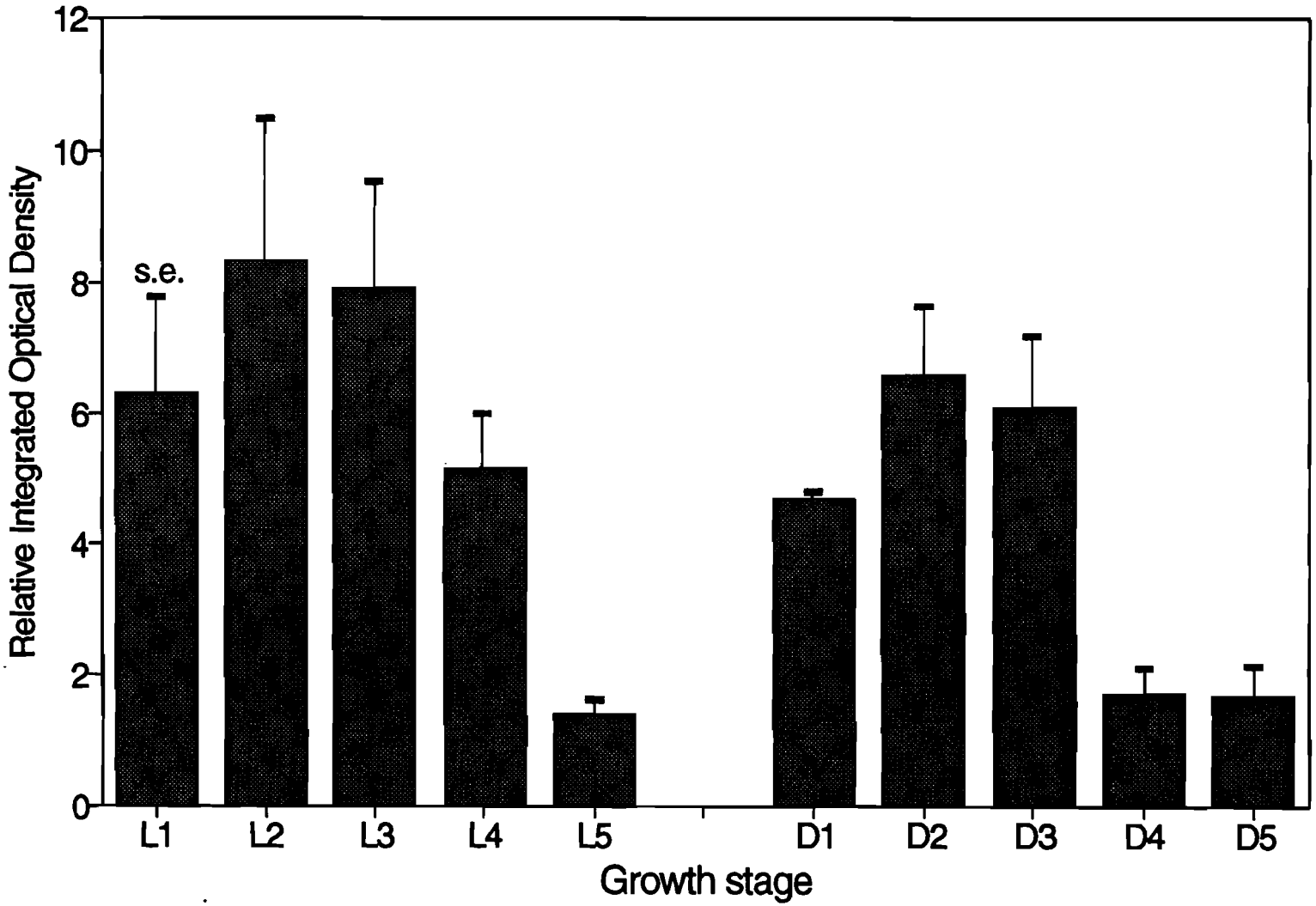

M
L1
L2
L3
L4
L5
D1
D2
D3
D4
D5
P
A
kb
4.40
2.37
1.35
B
2.37
1.35

Figure 3.17. Relative alliinase transcript abundance in developing *A. cepa* seedlings

Integrated optical density (IOD) values of signals produced by probing northern membranes of RNA extracted from *A. cepa* seedlings with the alliinase cDNA probe Alli6, were normalized relative to those obtained in the same lane by probing with a *Malus* 45S rDNA probe. For the purposes of graphical representation, the lowest corrected IOD value was assigned a value of 1 and all other values were calculated relative to this. Four replicate northern blots were analysed in this way and the average values are presented. s.e. - standard error.

Samples labelled 'L' are light-grown, those labelled 'D' are dark-grown.

Stage 1 = 1-2 days after imbibition (DAI)
Stage 2 = 2-3 DAI
Stage 3 = 3-6 DAI
Stage 4 = 6-9 DAI
Stage 5 = 10-14 DAI


12
10
8
6
4
2
0
s.e.
Relative Integrated Optical Density
L1
L2
L3
L4
L5
D1
D2
D3
D4
D5
Growth stage

Chapter 4. Discussion

4.1. Molecular cloning of alliinase

4.1.1. Screening of an *A. cepa* genomic clone bank with redundant oligonucleotides and cDNA probes

At the time this project was initiated. no molecular probes for alliinase-encoding sequences were available. Originally it was intended that a cDNA probe would become available from an *A. cepa* λgt10 library that was constructed and screened with oligonucleotide probes by J. Farrant (Crop and Food Research Ltd., Lincoln). To complement and extend this work, an *A. cepa* genomic library was constructed. However, no putative alliinase cDNA clones were isolated at this time (see Subsection 3.2). Therefore preliminary studies were conducted to establish the feasibility of screening the *A. cepa* genomic clone bank with redundant oligonucleotides designed from the 23 residue N-terminal alliinase peptide sequence, ALN1 (presented in Table 3.1).

Because the identity of the residue at position 11 was ambiguous and the two residues at the C-terminus were considered to be less certainly determined (Figure 2.1), the probes designed from ALN1 were limited to less than 30 nucleotides in length. This peptide sequence contained several amino acids with redundant codons (shown in Figure 2.1), incorporation of which would result in a highly degenerate oligonucleotide and a low abundance of the absolutely homologous probe. The strength of an authentic hybridization signal depends on the specific activity and the amount of the oligonucleotide bound to a target sequence. At degeneracies greater than 500-fold, and assuming realistic probe specific activities, specific hybridizations may be too infrequent to provide a useful signal (Wallace & Miyada 1987). For this reason the redundancies of the JL1 and JL2 probes were partially limited by making choices based on the codon usage analysis carried out by Murray *et al.* (1989), which indicates a preference for G or C in the third codon position of monocot genes. The redundancies of the JL1 and JL2 primers were limited to 384-fold and 84-fold respectively, to approximate probe complexities which have been successfully used in other studies. For instance, Weretilnyk & Hanson (1990) isolated cDNA clones of betaine-aldehyde dehydrogenase from a spinach library with two mixed 20mer probes of 128-fold and 96-fold degeneracy, and Whitehead *et al.* (1983) isolated target sequences from cDNA libraries with oligonucleotides of up to 384-

fold redundancy. To allow slightly mismatched hybrids to form, while also limiting the level of background binding, hybridization conditions were of moderate stringency (55°C with a 1x SSC, 0.1% SDS wash).

No positive clones were identified by this approach (see Subsection 3.2.2). The size of the *A. cepa* genomic clone bank (2.4 x 10^5 pfu) was small in comparison to the number of pfu generated using positive control ligation reactions (see Subsection 3.2.1) and was calculated to represent approximately 5% of the entire *A. cepa* genome. It is possible that this small clone bank size may have resulted from low ligation efficiencies during library construction, perhaps due to co-purification with the *A. cepa* DNA of an inhibitor of T4 DNA ligase. Contaminants such as polysaccharides and charged polymers are known to co-purify with plant DNA (Slightom & Drong 1988, Rogers & Bendich 1988). In addition, Graham *et al.* (1990) subsequently established that some *E. coli* host strains commonly used for genomic library construction, such as LE392, contain the loci *mcrA* and *mcrB*, which encode nucleases that cleave sites containing the modified base 5-methyl cytosine (m^5C). This modification is present in plant DNA at high levels (Ehrlich & Wang 1981), and a significant reduction in the size of plant genomic libraries constructed in LE392 has been demonstrated (Graham *et al.* 1990). This is consistent with the small size of the *A. cepa* genomic clone bank estimated by titration on LE392 (shown in Table 3.2 on page 87). As will be discussed in Subsection 4.2.2, Southern analysis later suggested that alliinase is encoded by a small gene family, hence this clone bank may not have contained alliinase genomic sequences. This hypothesis was later supported by the failure of an alliinase cDNA probe, Alli9, to form stable hybrids with genomic clones (see Subsection 3.6).

As discussed above, the high redundancy of the oligonucleotide probes JL1 and JL2 may have resulted in a low level of the absolutely homologous sequence being present in hybridization reactions. Subsequent DNA sequence analysis of an alliinase cDNA clone (discussed in Subsection 4.1.3) showed that this gene had a preference for A or T in the third base position (see Subsection 3.5.2.2), opposite to the monocot base composition bias indicated by the analysis of Murray *et al.* (1989). The JL1 sequence option most complementary to the alliinase cDNA contained two mismatched bases, while the JL2 sequence with the highest homology contained seven mismatches. The optimal calculated hybridization temperatures for these mismatched oligonucleotides are 44°C and 31°C respectively (Meinkoth & Wahle 1984). In retrospect it therefore appears that few stable authentic duplexes would be likely to form at the 55°C hybridization temperature used.

Several strategies have been devised to circumvent difficulties encountered using

redundant oligonucleotides for clone isolation. It has been demonstrated that longer probes (ie. more than 35 nucleotides) can hybridize stably even when there are several mismatched bases present (Kajimura *et al.* 1990). A longer oligonucleotide probe spanning the ambiguous residue at position 11 of the ALN1 peptide sequence may therefore have proved more successful. Alternatively, inosine, which forms weak bonds with any base, could have been used at positions where several choices were possible (Wallace & Miyada 1987), or additional probes could have been designed from other alliinase peptide sequences. Furthermore, the use of TMAC in hybridization buffers is known to remove the *T*m differences between A:T and G:C base pairs and thereby allows optimization of hybridization conditions to minimize non-specific probe annealing (Wozney 1990). In hindsight, the use of these strategies may have increased the possibility of identifying an authentic alliinase genomic sequence with redundant oligonucleotides. However, by this stage, other strategies in progress had identified several putative alliinase clones.

4.1.2. Preliminary studies on redundant oligonucleotide-primed PCR as an approach to cloning alliinase-encoding gene fragments

PCR is a powerful method for enrichment of specific target sequences (Mullis & Faloona 1987). At the time this work was initiated, many reports were available of successful gene amplifications using redundant oligonucleotide primers (eg. Touchet *et al.* 1987, Girgis *et al.* 1988, Gould *et al.* 1989, Gonzalez *et al.* 1989). Imperfectly matched primer/template duplexes are thought to be capable of priming product synthesis during transient binding as any extension immediately leads to an increase in their duplex stability (Gould *et al.* 1989).

Primers designed from alliinase peptide sequences ranged in length from 12 to 23 bases, resulting in oligonucleotides with 32- to 4096-fold redundancy. Amplifications were performed using a range of cDNA and genomic DNA templates, primers, nested primers, annealing temperatures and buffer conditions (summarized in Table 2.2). High levels of nonspecific, and/or spurious, amplifications were produced under most of the reaction conditions used (see Subsections 3.3.1 and 3.3.2), probably due to the high redundancy of the oligonucleotide primers. Bands of the sizes predicted for authentic alliinase amplification products were seldom observed.

The partial DNA sequence obtained from the termini of a single PCR product of approximately the size predicted for an alliinase transcript, revealed that it did not encode alliinase (Subsection 3.3.1.1, Figure 3.2 on page 95). This product appeared to result from two sequence variants of the redundant JL5 oligonucleotide acting to prime amplifications from non-alliinase sequences. Only a single transformant was subjected to DNA sequence analysis. In hindsight, further transformants should have been analysed, as a single PCR-amplified band may contain several products of the same size (Girgis *et al.* 1988).

The eventual availability of an alliinase cDNA clone (described in Subsection 3.5) revealed that four residues predicted by translation of the cDNA sequence, differed from those determined by direct peptide sequencing of native alliinase (see Subsection 3.5.2.1, page 106 and Figure 3.8). This resulted in the 3 bp at the 5' end of the JL12 primer, the 2 bp at the 3' end of the JL9 and JL13 primers, and the 3 bp at the 3' end of JL4, being non-complementary to the aligned alliinase cDNA sequence. However, the other redundant primers designed for PCR all contained a completely homologous sequence as one of the coding options. Alliinase protein was purified from *A. cepa* 'Pukekohe Longkeeper' whereas the mRNA for cDNA library construction was extracted from the cultivar Southport White Globe. It is therefore uncertain whether the four variant residues are due to inaccuracies of the peptide

sequence determination, the existence of alliinase isoforms, variation between cultivars, or a combination of these factors. If cultivar variation was responsible, these primers (JL4, JL9, JL12 and JL14) would be expected to be mismatched to the cDNA templates, which were synthesized from 'Southport White Globe' RNA, but may not be mismatched to the genomic DNA templates which were extracted from 'Pukekohe Longkeeper'. Binding at the 3' end is a major determinant in primer specificity (Sommer & Tautz 1989). Consequently, low homology of this region may in part explain the lack of alliinase sequence amplification with JL4, JL9 and JL13. Although binding of the primer at the 5' end is not as critical (McPherson *et al.* 1992b) it is conceivable that a lack of complementarity between the 5' end of the short primer JL12 and authentic target sequences might have lowered the stability of primer annealing.

The minimum requirement for specific amplification with short (ie. less than 30 base) primers has been defined as three perfectly matched bases at the 3' end (Sommer & Tautz 1989). One primer (JL12) designed from the alliinase peptide sequence achieved this, but only 12 bases of this primer were designed from the alliinase peptide sequence. The JL12 sequence option most similar to the subsequently isolated alliinase cDNA was predicted by the method of Binnie (1991), to occur 1016 times within the *A. cepa* genome by chance alone. Hence the probability of a single, authentic, product being amplified by this redundant oligonucleotide was low.

Re-amplification of products with an adjacent downstream sense primer ('nesting') (see Subsection 3.3.1.2) did not reduce the degree of background amplification, suggesting that the antisense primers, which were kept the same in both rounds of amplification, may anneal to multiple non-alliinase template sites. Haqqi *et al.* (1988) have demonstrated that increasing the dilution of first-round amplification products to 100 000-fold prior to re-amplification with a second set of primers can overcome amplification of spurious products in nested PCR experiments. This is presumably due to dilution of non-authentic amplification products present within the first-round reaction which could otherwise serve as templates during the second amplification round.

cDNA templates for PCR experiments were synthesized from *A. cepa* RNA isolated from root tip tissue (see Subsection 2.3.2.2). As no studies on alliinase transcript abundance in *A. cepa* root tissues have been performed, it is thought possible that difficulties in amplifying an alliinase-encoding sequence from cDNA may be accounted for by a low abundance of alliinase transcripts within root tip RNA. However this hypothesis could not be supported by the data generated by this

study, and there is no published evidence to defend or dismiss this hypothesis.

Various approaches have been published to improve the specificity of PCR with redundant oligonucleotides. These include the use of inosine (Knoth *et al.* 1988, Patil & Dekker 1990), or only G and T (Touchet *et al.* 1987), at redundant third base positions. Hot starts (Chou *et al.* 1992), optimization of KCl concentration (Saiki 1989), addition of formamide (Sarkar *et al.* 1990), or the use of TMAC (Hung *et al.* 1990), have also been shown to improve amplification specificity. Alternately, sets of less redundant primers could have been synthesized containing fewer mismatches at their 3' end, or from further alliinase peptide sequence data. It was evident from these preliminary investigations, however, that with the limited amino acid sequence data available, the probability of successfully amplifying an alliinase-specific sequence by PCR was low, and moreover might require extensive studies to determine optimal conditions and/or involve multiple primer options. Furthermore, immunoscreening of an *A. cepa* expression library had by this time resulted in the detection of several putative alliinase clones. These are discussed in the next section.

4.1.3. Isolation of alliinase cDNA clones by immunoscreening of an *A. cepa* cDNA expression library

4.1.3.1. Relationship of the clones

Seven immunopositive clones were isolated from an *A. cepa* cDNA expression library. Of these, one (Alli8) was not DNA sequenced but contained identical *Xho* I and *Eco* RI restriction sites to another clone that was sequenced (Alli4B). The cDNA clones Alli6 and Alli4B were sequenced entirely, whereas the termini of the clones Alli7, 4b and 9 were sequenced (described in Subsection 3.5.2, and illustrated in Figure 3.7). The rationale for this was based on the observation that in a number of cases, transcribed sequences of gene family members showed maximum divergence in their untranslated regions (Dean *et al.* 1985). Due to time limitations none of the clones were sequenced entirely on both strands. Double stranded, continuous, sequencing is normally a requirement to ensure the DNA sequence obtained is accurate. However, as shown in Figure 3.8, alliinase peptide sequences aligned to the longest open reading frame of the clones at several sites, indicating that the sequence data was accurate throughout. The DNA sequences determined from the five clones Alli4B, 6, 7, 8B and 9 were almost identical, with the only variation found between four of these sequences being the length of the 5' end. This indicated that these cDNAs are most likely to be derived from the same gene. The fifth clone (Alli4B) was 37 bases shorter at the 3' end with six different bases just prior to the polyadenylate tail, suggesting that this transcript may have arisen from a recently diverged gene, an allelic variant, or a different post-transcriptional modification. The possibility that this variation may be a result of a cloning artifact could not be dismissed, but the presence of an intact polyadenylate tail suggested this was unlikely. This six base pair difference was the only sequence variation found between these five similar cDNAs, however their independence is confirmed by the observation that they all vary in length either at the 5' or the 3' end.

By the time the DNA sequencing of these clones was completed, four regions of alliinase amino acid sequence had been ascertained. The homology between the cDNA-inferred and directly-determined alliinase peptide sequences was high, with 55 of the 59 amino acid residues determined by peptide sequencing aligning with the protein sequence derived from the longest ORF of the cDNA clones. As mentioned previously (see Subsection 4.1.2), four differences were found between the amino acid sequence predicted from the cDNA sequence and that determined from alliinase protein. These may be explained by difficulties in the protein sequence determination, cultivar variation, or polymorphism. The uniqueness of the determined alliinase cDNA sequence and the derived alliinase peptide sequence is demonstrated by their lack of homology to other sequences in the DNA or protein

sequence databases (see Subsection 3.5.2.4).

A further immunoreactive clone, Alli4A, was also isolated and fully DNA sequenced. This clone did not contain sequences related to alliinase nor to any DNA sequences within the GenBank database (see Subsection 3.5.2.4 and Appendix I). The size of the polypeptide encoded by the longest ORF was 24.2 kDa, however as no stop codons were present upstream of the most 5' methionine it could not be unequivocally determined if the entire longest ORF was encoded.

A. cepa alliinase is believed to be composed of three (Tobkin & Mazelis 1979) or four (Nock & Mazelis 1987, Jansen *et al.* 1989a) subunits, each of about 50 kDa. Multiple bands have been detected by isoelectric focusing, and this was attributed to heterogeneity within the subunit oligosaccharide side chains (Nock & Mazelis 1987). However Jansen *et al.* (1989a), using a different purification procedure, demonstrated that the *A. sativum* alliinase migrated as a single isoform. Therefore it is improbable that the Alli4A sequence encodes an uncharacterized alliinase subunit. It is more likely that the protein expressed by the λAlli4A clone contained an epitope to which the anti-alliinase polyclonal antibodies were cross-reactive. Furthermore, the alliinase preparation used to induce anti-alliinase antibodies was not purified by SDS-PAGE prior to antibody induction, thus the possibility that it may have contained low levels of contaminating immunogenic *A. cepa* proteins cannot be dismissed. A cross-reactive protein of 24.2 kDa could not, however, be detected by western blotting of alliinase preparations with anti-alliinase polyclonal sera (M.L. Shaw pers. comm.).

4.1.3.2. Nucleotide sequence analysis

4.1.3.2.1. Polyadenylation signals and sites

The sequences of the alliinase clones Alli6, Alli7, Alli8B and Alli9 all contained three potential polyadenylation signals. The Alli4B cDNA had a truncated 3' untranslated region and consequently lacked the third polyadenylation signal. The upstream polyadenylation signal lies 23 base pairs 5' to the polyadenylate tail of Alli4B, while the downstream polyadenylation signal lies 13 bases 5' to the poly(A) tail of the other alliinase cDNA sequences (Figure 3.8, page 121). Multiple polyadenylation signals and corresponding polyadenylation addition sites are frequently found in plant transcripts (Messing *et al.* 1986, Dean *et al.* 1986), for example, some cDNAs encoding patatin in potato contain up to five potential

polyadenylation signals (Mignery *et al.* 1988), and Koes *et al.* (1986) attributed the nine different poly(A) tail start sites found among a number of petunia CHS cDNAs to the presence of three potential polyadenylation signals. Variation in the usage of polyadenylation signals may result from differences in the efficiencies of recognition of the signal and its flanking sequences (Ingelbrecht *et al.* 1989), or to the presence of a number of upstream *cis* elements regulating polyadenylation at various sites (Mogen *et al.* 1992), and it has been postulated that differences in the 3' tails of transcripts may lead to differential mRNA stability (Dean *et al.* 1985).

4.1.3.2.2. ORFs and leader sequences

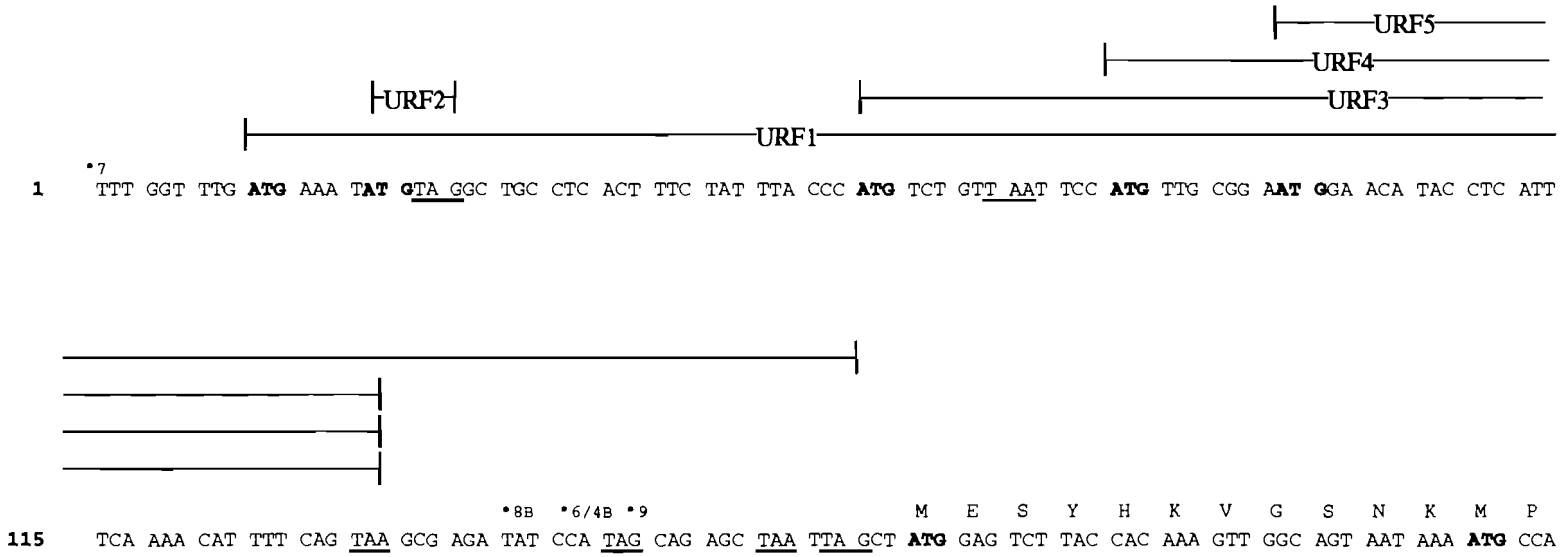
The longest ORF in the alliinase cDNA sequence began at position 136 relative to the Alli7 sequence and terminated at a stop codon at position 1573. The Alli7 cDNA contained an upstream leader sequence 111 bases longer than the other alliinase cDNAs, which could be accounted for by the extent of cDNA synthesis at the 5' end. This 135 nucleotide region upstream of the longest ORF contained several possible short ORFs (see Subsection 3.5.2.1, and Figure 4.1, page 159). The other alliinase cDNA clones, Alli4B, 6, 8B and 9 began between positions 112 and 120 (Figure 3.8). The longest ORF appeared to encode the entire alliinase polypeptide sequence as it was preceded in all the alliinase cDNA clones by stop codons at positions 127 and 131.

The scanning model predicts that eukaryotic mRNAs are translated by ribosomal initiation at the most 5' AUG of a transcript if it lies in a favourable context (Kozak 1981). Furthermore, the most 5' AUG of plant genes is usually the start of the longest ORF (Joshi 1987a). However the presence of upstream reading frames (URFs) in some eukaryotic genes suggests that other means of translation initiation may occur. Current models of eukaryotic translation suggest initiation factors usually dissociate from the ribosome upon each initiation event (eg. at the URF start codon) and are no longer available for further downstream translation (Wahba *et al.* 1990). However studies in animal systems have demonstrated resumption of scanning after translation of an URF and recovery of initiation competence, possibly by binding of a new set of initiation factors. The efficiency of this reinitiation has been found to vary and is dependent on the distance between the upstream stop codon and the downstream initiation site (Kozak 1987).

Figure 4.1. Upstream reading frames (URFs) present within the 5' leader sequence of Alli7

The 5' ends of the various cDNA clones (Alli4B, 6, 7, 8B and 9) are indicated by a ● and the corresponding number.

ATG - possible translation initiation start codons.
TAA, TAG - possible translation termination codons.


URF5
URF4
URF3
URF2
URF1
•7
1 TTT GGT TTG ATG AAA TAT GTA GGC TGC CTC ACT TTC TAT TTA CCC ATG TCT GTT AAT TCC ATG TTG CGG AAT GGA ACA TAC CTC ATT 87
•8B •6/4B •9
M E S Y H K V G S N K M P
115 TCA AAA CAT TTT CAG TAA GCG AGA TAT CCA TAG CAG AGC TAA TTA GCT ATG GAG TCT TAC CAC AAA GTT GGC AGT AAT AAA ATG CCA 174

It has been postulated that regulation of dicot phytochrome gene expression may be partly at the level of translation as transcripts encoding phytochrome in zucchini (*Cucurbita* sp.) (Sharrock *et al.* 1986), pea (Sato 1988), and *Arabidopsis* (Sharrock & Quail 1989) contain at least one URF (Sharrock & Quail 1989), although they are notably absent from the corresponding genes in monocots (Christensen & Quail 1989, Hershey *et al.* 1985, Kay *et al.* 1989). A recent study using *Orychophragmus violaceus* protoplasts has shown an URF consisting of a start codon immediately followed by a stop codon inhibited translation from a downstream ORF encoding a GUS marker gene to 50% and a 100-codon URF to almost undetectable levels, indicating the effect of URFs on protein synthesis may be similar in plant and mammalian systems (Fütterer & Hohn 1992). The presence of possible URFs in the 5' region of Alli7 should similarly affect alliinase translation. The presence of multiple termination codons just prior to the start of the longest ORF of the alliinase cDNAs suggests that ribosomal reinitiation would then be required for translation of the alliinase transcript.

To ascertain which of the various methionine codons within the Alli7 cDNA sequence are most likely to act as translation initiators, comparisons were made to consensus sequences of translation initiation flanking sequences. The sequence CCRCCAUGG, emerged from a survey of 699 vertebrate mRNAs as the most favourable consensus sequence for initiation in higher eukaryotes (Kozak 1984). The most highly conserved nucleotides in the above consensus sequence occurred at -3 and +4 relative to the AUG codon. The -3 position must be a purine and the +4 is a G. However compilation analyses have suggested that the selection of the AUG initiator differs in plants and animals. Lütcke *et al.* (1987) and Joshi (1987a) independently identified the sequence AACAAUGGC as the most favoured initiator sequence in plants. Furthermore the consensus sequence for monocot and dicot plants has also been shown to vary, with GCARCCAUGGC found to be the consensus start site from 93 monocot genes, and AAAAAMAUGGC the consensus arrived at from analysis of 233 dicot genes (Cavener & Ray 1991). In all these analyses the most consistent feature is the presence of A or G nucleotides in the -3 position, and a G in the +4 position. Comparison of the Alli7 ATG codons and flanking regions to the monocot consensus sequence from Cavener & Ray (1991) indicates that the ATG at position 46, and that present in all the alliinase cDNAs at position 136, are in the most favoured sequence contexts as both of these sites contain G in the -3 and +4 positions. The start sites at positions 10 and 46 are therefore the most likely to result in the translation of a protein from the Alli7 leader sequence, based on the scanning model and flanking sequences respectively.

Termination of translation initiated from these methionines would be most likely to occur at the stop codon at position 103 (see Figure 4.1), resulting in a protein of either 3643 Da or 2227 Da. However, as discussed by Cavener & Ray (1991), the concepts of consensus and optimal are not equivalent, and many exceptions to consensus flanking sequences can also function as favourable contexts for initiation of translation.

The unusual structure of this 5' leader sequence, and its presence in only one of the five independently isolated alliinase clones sequenced, suggests that it may be an artifact of the cloning procedure. Isolation and sequencing of genomic alliinase clones, and the identification of the start of alliinase transcripts by primer extension or ribonuclease protection studies, will be of value in determining its authenticity.

4.1.3.2.3. Codon bias and base composition

Higher plants exhibit unique codon biases, with plants of the same taxonomic class maintaining a similar codon usage pattern (Campbell & Gowri 1990). Variation mainly occurs in choices between codons ending in G or C *versus* those ending in A or T. Strong differences have been found between the codon usage of nuclear plant genes from monocot and dicot plants, with monocot genes having a predominance of G or C in the third base position (Murray *et al.* 1989). Campbell & Gowri (1990) found that on average 45% of dicot gene codons, and 73.5% of monocot gene codons, ended with C or G. These authors also found, however, that the base composition of monocots exhibited a bimodal distribution, with some genes showing a broader codon usage and a lower percentage of codons with G or C as the third base.

About 40% of alliinase codons in the longest ORF used G or C in the third base position (see Subsection 3.5.2.2), which, by comparison to the majority of monocot genes sequenced to date, is low, and is more similar to the composition of dicot genes. This low G or C content is also more consistent with that of total *A. cepa* DNA which has been estimated at 34.6% G+C (Kirk *et al.* 1970). Codon usage analysis of monocot genes has been dominated by a disproportionately high representation of sequences from maize and other cereals (Campbell & Gowri 1990). Consequently, the observed disparity between the alliinase cDNA base composition and the average values derived by compilation analysis of monocot genes is not surprising given the substantial taxonomic distance of the *Alliaceae* from the *Gramineae* (Dahlgren & Bremer 1985, Chase & Palmer 1989).

4.1.3.3. Features of the derived alliinase protein sequence

To provide further evidence that the cDNA clones encode alliinase, the peptide sequence derived by translation of the longest ORF was analysed for features consistent with those reported for the purified *A. cepa* alliinase protein. It was anticipated this analysis might additionally produce further information on the structure and function of this enzyme.

4.1.3.3.1. Signal peptide structure

Vacuolar proteins reach their final location after translocation through the endoplasmic reticulum (ER) and the golgi apparatus (Tague and Chrispeels 1987). The alliinase polypeptide sequence predicted from the cDNA clones contained a hydrophobic N-terminal region terminating at a peptidase cleavage site motif between residue 32 and 34 (Perlman & Halvorsen 1983)(see Subsection 3.5.2.3 and Figure 3.10). Given that alliinase is located in the vacuole (Lancaster & Collin 1981), it is most likely that this prepeptide functions as an ER targeting signal. The most favoured site for cleavage of this region was predicted by the statistical approach of von Heijne (1986), to lie between residues 27 and 28. The site between residues 34 and 35 was assigned as the second most favoured. The predictive accuracy of this calculation has been estimated at approximately 78%, with 19 of 161 known cleavage sites being assigned the second highest score (von Heijne 1986). The alignment of the mature alliinase N-terminus at residue 35 indicates that the most probable site of cleavage is after residue 34. A signal peptide length of 34 residues is atypical in eukaryotic proteins, although it is not outside the range previously found (Pugsley 1989). Equally long ER targeting peptides have been reported for other plant prepeptides, such as the 34 amino acid prepeptides of the lectins of *Ricinus communis* (Tregear & Roberts 1992).

Vacuolar or protein storage body targeting information resides within the N- or C-terminal regions of some plant protein precursors. This has been demonstrated conclusively with the sweet potato protein sporamin, in which the preproprotein contains a 21 amino acid-long signal sequence followed by a 16 residue propeptide which is sufficient for vacuolar sorting (Matsuoka & Nakamura 1991). Domains functional in vacuolar targeting have also been identified at the C-terminus of the barley lectin protein (Bednarek & Raikhel 1991) and between amino acids 33 and 44 of bean phytohemagglutinin (Tague *et al.* 1990), but these show little or no amino acid sequence conservation. Efficient vacuolar targeting of the plant thiol protease proaleurin has been shown to be mediated by the combined action of at least three

five- and eight-residue contiguous determinants located near the N-terminus, two of which are sufficient to target chimeric proteins in transgenic tobacco protoplasts (Holwerda *et al.* 1992). A variety of sequences or protein motifs such as β-sheets and loops have been implicated in vacuolar targeting (Sebastiani *et al.* 1991) and may be involved in presenting the vacuolar sorting determinants in an appropriate context (Chrispeels & Raikhel 1992).

The C-terminal extension of barley lectin responsible for vacuolar targeting is predicted to form an α-helical structure (Wilkins & Raikhel 1989), and an α-helical structure is also predicted for the N-terminus of the mature alliinase subunit (Figure 3.11). However other plant vacuolar targeting signals do not consistently show this propensity, and the only general consensus among the known determinants is that they contain a number of residues with hydroxyl groups (serine or threonine residues) interspersed with both negatively and positively charged residues (Holwerda *et al.* 1992). As the predicted nascent alliinase protein does not contain any sequence motifs analogous to those shown to function in targeting of other plant vacuolar polypeptides, it was not possible to predict which regions of the inferred protein sequence are most likely to fulfil this role.

4.1.3.3.2. Predicted size of the alliinase subunit protein

The largest immunoreactive peptide expressed from the pAlli6 cDNA by *E. coli* migrated at 47 kDa on SDS-PAGE. This compares favourably with the 47 kDa size estimated for mature, deglycosylated alliinase subunits purified from *A. cepa* bulbs (M.L Shaw pers. comm.) but not with the 54.884 kDa size of alliinase subunits inferred from the longest ORF of the cDNA clones. *E. coli* would be expected to cleave the alliinase precursor at the signal peptidase recognition site between residues 34 and 35 (Perlman & Halvorsen 1983), which would result in a predicted alliinase protein of 51.135 kDa. This 4 kDa difference in size together with the presence of a second immunoreactive species at 41.4 kDa suggests that alliinase may be cleaved, or that translation is initiated from internal methionine residues in *E. coli*. Translation from the second methionine residue present in the mature predicted protein sequence at position 124 results in a calculated size of 41.54 kDa, suggesting this may be the explanation for the smaller immunoreactive band. The ATG encoding this methionine is preceded 7 bp previously by the DNA sequence motif 'AGGATGG' which is similar to the Shine-Delgarno sequence 'AGGAGGU'. However, no methionine codons were present in the predicted alliinase amino acid sequence which could account for the 47 kDa band. The size of this immunoreactive

protein may be due to a post-translational modification, proteolytic cleavage, or anomalous migration of the *E. coli* translation product on SDS-PAGE. Other studies in which cloned cDNAs were expressed in *E. coli* have also produced proteins with migrations different from that predicted. Reynolds *et al.* (1992) found that a cDNA encoding aspartate aminotransferase-P_2 isolated from lupin root nodules, encoded a protein with a predicted size of 50 kDa but three immunoreactive peptides of 55 kDa, 39 kDa and a doublet of 43/45 kDa were expressed in *E. coli* within the pBluescript(-) plasmid. These various translation products were attributed to initiation at methionine codons both upstream and downstream of the start of the cDNA sequence.

4.1.3.3.3. Protein sequence analysis

Hydropathy analysis of derived proteins can be used to predict if a protein is most likely to be soluble or membrane-bound. The portions of membrane-bound proteins that are located within the lipid bilayer are clearly delineated by large uninterrupted hydrophobic areas (Kyte & Doolittle 1982). This approach has also been used to accurately predict internal and external regions of the globular proteins bovine chymotrypsinogen and dogfish lactate dehydrogenase (Kyte & Doolittle 1982). The predicted mature alliinase protein did not contain any hydrophobic portions of sufficient length to infer a membrane spanning region (Klein *et al.* 1985). The hydropathy profile obtained (see Figure 3.10) is consistent with that of a soluble, globular protein and suggests alliinase is free within the vacuole. The hydrophobic areas between residues 160 to 162 and 327 to 330 would be predicted to locate toward the interior of the protein, while the hydrophilic areas between 101 to 108, 197 to 216 and 309 to 321 would be more likely to be exposed to the external aqueous environment. A further analysis which calculates regions of most likely antigenicity (Welling *et al.* 1985) predicted multiple immunogenic sites located throughout the protein (see Subsection 3.5.2.3). This regular distribution of epitopes suggests that clones containing partial sequences would express proteins with fewer epitopes, which is compatible with the consistent isolation from the *A. cepa* cDNA expression library of clones containing entire ORFs (see Subsection 3.5.2.1).

It is not known if alliinase contains disulphide bridges, but the inhibition of disulphide bond formation has been postulated as a requirement for the maintenance of alliinase activity. As the sulphydryl reagent *N*-ethylmaleimide has no affect on alliinase activity, this does not appear to be associated with the maintenance of a sulphydryl group (Tobkin and Mazelis 1979). The inferred mature alliinase protein

contained 11 cysteine residues, and a twelfth is located within the predicted ER targeting peptide (see Subsection 3.5.2.3). Five of these are clustered within 19 residues in a region of predicted β-turns between residues 53 and 100 and two further cysteine residues were also predicted to lie in close proximity within a predicted β-turn region at amino acid positions 405 and 413. The other five predicted cysteine residues are spread throughout the peptide sequence. This concentration of cysteine residues within predicted multiple turn regions implies the formation of loop structures linked by disulphide bridges and suggests these regions may have specific functions in alliinase enzyme structure or activity. However no conclusions can be drawn from this data, and protein structural analyses such as x-ray crystallography are necessary to define the role of these cysteine residues.

4.1.3.3.4. Glycosylation of alliinase

Vacuolar proteins, such as alliinase, are believed to transit the intracellular endomembrane system (Tague & Chrispeels 1987). Glycans with the structure $(Glucose)_3$-$(Mannose)_9$-$(N$-acetylglucosamine$)_2$ are attached to Asn-X_{aa}-Ser/Thr sites as proteins are cotranslationally secreted into the ER (Kornfeld & Kornfeld 1985). These high mannose side-chains can be further modified within the golgi of plant cells, to form complex glycans containing fucose and xylose in addition to *N*-acetylglucosamine and mannose (Faye & Chrispeels 1988).

The presence within the predicted alliinase protein sequence of Asn-X_{aa}-Ser/Thr motifs (see Subsection 3.5.2.3) is compatible with the known glycosylated nature of the *A. cepa* enzyme (Tobkin & Mazelis 1979). Oligosaccharides have been estimated to represent between 4.6% (Nock & Mazelis 1987) and 5.8% (Nock & Mazelis 1986) of the weight of *A. cepa* alliinase, consistent with the 3 kDa size decrease found after deglycosylation with TFMS (see Subsection 3.4.2, page 98). Glycan sidechains can vary greatly in size with a minimum of approximately 1.3 kDa (Goodwin & Mercer 1983). Consequently a size change of 3 kDa could be accounted for by the removal of two small oligosaccharide side-chains or a single, larger, side-chain. Whatever the case, it is clearly unlikely that all four sites present within the predicted alliinase protein sequence are used.

On average, only one-third of potential Asn-linked glycosylation sites in proteins are filled (Kornfeld & Kornfeld 1985). A site is more likely to function if it resides within a β-turn (Kornfeld & Kornfeld 1985), and is situated toward the N-terminus of the protein (Gavel & von Heijne 1990). Two different secondary structure

predictions based on the deduced alliinase amino acid sequence consistently located both the N-terminal Asn-X_{aa}-Ser/Thr sites at amino acid positions 53 and 180 within β-turn regions (see Subsection 3.5.2.3 and Figure 3.11). The glycosylation site predicted toward the C-terminus of alliinase at position 362 was estimated by both predictive approaches to be near, but not within, a β-turn region. This evidence suggests that Asn-linked glycosylation is more likely to occur at the most N-terminal site of alliinase, and may also occur at the site at position 180.

4.1.3.3.5. PLP cofactor binding

The sequence motif Ser/Thr-X_{aa}-X_{aa}-Lys is the site of attachment of PLP in a number of diverse PLP-requiring enzymes, such as aminocyclopropane-1-carboxylate synthase (ACC) of tomato (Yip *et al.* 1990) and L-histidinol phosphate aminotransferase from *Salmonella typhimurium* (Hsu *et al.* 1989). PLP is thought to bind to the lysine residue, while the 5' phosphate of the PLP hydrogen bonds to the hydroxyl residue of the serine or threonine (Ford *et al.* 1980, Tanase *et al.* 1979). Four of these conserved sequences are present within the polypeptide derived from the cDNA clones (see Subsection 3.5.2.3 and Figure 3.8), suggesting the alliinase PLP cofactor may also bind to this motif. One mole of PLP has been found to bind to each alliinase subunit (Tobkin & Mazelis 1979, Nock & Mazelis 1987). Therefore only one of the four sites present in the predicted alliinase protein sequence may be involved in PLP binding. Many eukaryotic and prokaryotic aminotransferases show a similar PLP-binding site context, and flanking sequences may have a role in determining a functional active site. Yip *et al.* (1990) reported 50-60% sequence homology between the peptides surrounding the active sites of the apple and tomato ACC synthases and those from cytosolic and mitochondrial aspartate aminotransferases, as well as histidinol-phosphate aminotransferases from yeast and *S. typhimurium*. However, comparisons of the potential PLP binding sites of alliinase and those of other PLP-requiring enzymes compiled within the PROSITE database (Bairoch 1992) did not establish any substantial sequence conservation however, indicating that alliinase uses a novel PLP-binding site context.

4.2. Estimation of the number of genes encoding alliinase in the *A. cepa* genome

4.2.1. Restriction endonuclease digestion of *A. cepa* DNA

Plant DNAs are highly methylated with the level of m^5C in some species amounting to approximately 30% of the total cytosines. Methylated bases are usually distributed between the sequences m^5CpG and m^5CpXpG (Antequera & Bird 1988). *A. cepa* DNA was recalcitrant to fragmentation with the enzymes *Cla* I, *Nru* I, *Pst* I, *Sal* I and *Xho* I, all of which contain the methylation sequences CpG or CpXpG within their recognition sites (see Subsection 3.7). These motifs were absent from the recognition sites of enzymes which significantly fragmented *A. cepa* DNA and consequently site methylation was thought to be a possible cause for the differential extent of cleavage observed. Alternatively, as the doublet CpG is known to be avoided in plant nuclear sequences (Boudraa & Perrin 1987), appropriate G+C-rich recognition sites may be under-represented, particularly within the A+T-rich nuclear genome of *A. cepa* (Kirk *et al.* 1970). Enzymes with recognition sites lacking CpG or CpXpG motifs were subsequently used to digest *A. cepa* DNA for Southern analysis.

4.2.2. Representation of alliinase-encoding sequences in *A. cepa*

To ascertain if a single gene or a gene family encodes alliinase, Southern blots of *A. cepa* (cultivar Pukekohe Longkeeper) genomic DNA were probed with the representative cDNA Alli6. Southern analyses were initially performed under conditions of high stringency (0.1x SSC, 65°C wash). All of the seven restriction enzymes employed in *A. cepa* Southern analysis resulted in more than one hybridizing fragment (see Subsection 3.7, Figures 3.13 and 3.14). The multiple hybridizing fragments observed in all lanes indicated the probable existence of an *A. cepa* alliinase gene family.

4.2.2.1. Comparison of hybridization intensities to single copy reconstructions

Gene copy number estimations cannot simply be derived from the number of bands produced by hybridization of a cDNA sequence to restriction endonuclease-

fragmented genomic DNA, as multiple fragments can be produced from a single gene if sites for a restriction enzyme reside within the transcribed sequence or within the intron regions of the corresponding gene. In non-haploid tissues, multiple bands can arise from both allelic variation at a single locus and from discrete homologous loci within the genome.

Hybridization signal intensity can vary, depending on the degree of overall sequence similarity or the presence of multiple or partial genomic sequences within a particular fragment. Hybridization signals of high intensity can result from clustering of gene family members on the same restriction fragment. The comigration of two independently hybridizing fragments on agarose gels is also possible. The unequivocal determination of the existence of multiple gene copies therefore requires isolation of full genomic clones. However estimates of gene copy number can be made by comparing the hybridization intensity of fragments to a control lane containing the DNA probe at a concentration calculated to represent a single copy of the genomic sequence.

The hybridization intensity of bands produced by Southern analysis of *A. cepa* DNA were compared to single copy reconstructions using an alliinase cDNA probe, Alli6 (see Subsection 3.7). Many bands hybridized at intensities lower than single copy, indicating these fragments either contained a partial homologous sequence, or a sequence of lower overall homology. This was particularly clear with *A. cepa* DNA digested with *Hind* III and *Eco* RV, enzymes for which cleavage sites are predicted within the alliinase cDNA sequence at positions 1264 and 964 respectively (Figure 3.9, page 123). The *Eco* RV digestion resulted in eight bands of intensity approximately half that of the single copy reconstruction. There was also a band of greater than single copy intensity, and another that was very faint. Digestion with *Hind* III produced ten bands, four of which showed strong hybridizations, and six which were fainter. *Xba* I and *Eco* RI sites are located toward the 5' end of the alliinase cDNA sequence at positions 440 and 326 respectively. *Xba* I-digestion produced three strong bands, as well as one that was faint, while *Eco* RI-digestion produced five strongly hybridizing bands. The patterns produced by cleavage of *A. cepa* DNA with *Hind* III, *Eco* RV, *Xba* I and *Eco* RI could result from multiple cleavage sites both within the coding and non-coding regions of homologous genomic sequences. However these patterns are unlikely to have been generated from a single fragmented gene as the summed hybridization intensities were in each case considerably greater than that of a single genome equivalent, suggesting that more than three alliinase genes are present in the *A. cepa* genome.

Although no *Dra* I sites were predicted within the alliinase cDNA sequence, this restriction endonuclease produced eight DNA fragments with homology to the Alli6 probe. Three of these were of low binding intensity and were smaller than the 1.6 kb Alli6 cDNA, indicating that they did not contain intact homologous sequences. A polymorphism was apparent in the patterns produced by this enzyme from two *A. cepa* bulbs. *A. cepa* is generally regarded as an outcrossing species. Although there appears to be no incompatibility system, individual flowers on an umbel are highly protandrous, effectively preventing self-pollination of a single flower (Currah & Ockendon 1978). Therefore genetic heterogeneity within a population is expected. DNA isolated from the two *A. cepa* bulbs also resulted in slightly different hybridization patterns when digested with *Eco* RI. Variation in the extent of cleavage by certain restriction endonucleases such as *Eco* RI is well documented, particularly over long digestion periods (Brooks 1987). Therefore it is not clear if the variable hybridization patterns produced by this enzyme are due to differential cleavage or to a further polymorphism.

Enzymes for which no sites were predicted within the alliinase cDNA clone (*Bgl* II and *Bam* HI) generally produced few hybridizing bands (five and three respectively). Four of the bands produced by *Bgl* II were about equivalent in intensity to single copy approximations. Two of the bands produced by *Bam* HI-digested *A. cepa* DNA hybridized with greater intensity than single copy reconstructions, suggesting either co-migration of fragments containing homologous sequences, or clustering of some alliinase genes. Clustering would suggest that originating gene duplication events were relatively recent (Li *et al.* 1985).

An accurate estimate of the number of copies of the alliinase gene within the *A. cepa* genome is difficult to make from this Southern analysis data, as variable numbers of bands and a range of hybridization intensities were produced by the different restriction enzymes employed. However, the intensity of hybridization from all digestions was consistently greater than that of three genome equivalents per haploid genome, suggesting that alliinase is encoded by a gene family of at least four members. The largest number of bands produced was ten, but the hybridization intensities of these indicated that they were unlikely to contain entire homologous sequences. These hybridizations were carried out under high stringency conditions (0.1x SSC wash, 65°C), indicating the multiple alliinase genomic sequences detected have considerable sequence correspondence. This estimate is probably a minimum one because functionally related DNA sequences that are distantly related would not have been detected under the hybridization conditions used. Furthermore, genes of

lower sequence similarity may hybridize at lesser intensity than a gene fully complementary to the Alli6 probe, and therefore result in an underestimate of gene copy number. Consequently, Southern analysis was performed under more permissive conditions, ie. a 2x SSC wash at 65°C, to ascertain whether additional alliinase-related sequences were present within the *A. cepa* genome.

Under these conditions no further *Bgl* II, *Bam* HI or *Xba* I-digested *A. cepa* DNA fragments hybridized to the cDNA probe Alli6 (see Subsection 3.7, Figure 3.15). One extra band was detected within DNA digested with *Dra* I. Several additional bands were detected from *Eco* RI-digested DNA. However, as discussed above, these may be due to secondary endonucleolytic activity rather than to the presence of further alliinase-related sequences. Thus *A. cepa* alliinase appears to be encoded by a small gene family.

4.2.2.3. Hybridization analysis with an untranslated region probe

Sequence divergence of related genes is a reflection of the opposing effects of nucleotide substitutions and functional constraint, and hence can be used to study the evolutionary relationships of gene family members (Li *et al.* 1985). This can be determined by altering the hybridization stringency of Southern blots or by using short probes specific to regions of differing functional importance. UTR sequences of mRNAs tend to show significant divergence between related genes (Dean *et al.* 1985) and this variability can be used to discriminate between gene family members. For example *cab* cDNA 3' UTR probes from petunia hybridized to specific subsets of fragments on genomic Southern analysis, even under relatively low stringency hybridization conditions (Dunsmuir *et al.* 1983). Furthermore the probability of a short DNA probe hybridizing to two exons is low, and therefore can more accurately determine gene copy number. For these reasons Southern hybridization was performed with the 276 bp probe Alli6-3', which contained the entire 3' UTR of the alliinase cDNA clone Alli6. No sites for the restriction enzymes used for Southern analysis reside within the 3' UTR of Alli6. Thus, except in the event of uncharacterized alliinase sequences possessing an additional restriction endonuclease site within the small region spanned by this probe, Alli6-3' would be expected to hybridize to only one fragment per alliinase gene or pseudogene.

For each restriction endonuclease tested, Alli6-3' hybridized most strongly to a single fragment (see Subsection 3.7 and Figure 3.13), suggesting that one *A. cepa* alliinase gene is of highest homology to the cDNA clones isolated in this study.

That the alliinase cDNAs isolated are apparently derived from a single, or two very closely related, gene(s) is consistent with their high sequence similarity. In each lane between three and four other fragments previously identified by Alli6 also bound to Alli6-3' at lower intensity, in agreement with the presence of several *A. cepa* genomic sequences of lower homology to the Alli6 cDNA and indicative of sequence divergence among the members of the gene family. Hence *A. cepa* alliinase appears to be encoded by a small gene family with a minimum of three or four members, one of which is of highest homology to the cDNAs isolated in this study.

Heterogeneity of *A. cepa* alliinase has been detected by isoelectric focusing, but was ascribed to carbohydrate sidechain variation (Nock & Mazelis 1987), and no isoforms of alliinase subunits have yet been described. The presence of several alliinase genes with divergent sequence homologies suggests that isoforms may exist. However gene family members can show considerable differences at the nucleotide sequence level and still encode products of high amino acid sequence similarity, as has been demonstrated among the *rbcS* gene products of several plants (Dean *et al.* 1989b).

The isolation from the *A. cepa* library of five discrete cDNA clones which appear to be derived from the same, or possibly two closely related, gene(s) suggests that some members of the alliinase gene family detected by Southern analysis are not transcriptionally active in sprouting bulb shoot tissue, or are transcribed at lower levels. Differential expression of plant gene family members in specific plant tissues, at different developmental stages or in response to environmental stimuli is well documented. This has been clearly demonstrated with the light regulated genes involved in photosynthesis (eg. Dean *et al.* 1985, Simpson *et al.* 1986b), the chalcone synthase genes (Koes *et al.* 1986, Ryder *et al.* 1987, Block *et al.* 1990) and the actin gene family members of soybean, petunia and rice (Hightower & Meagher 1985, Baird & Meagher 1987, McElroy *et al.* 1990a). A further possibility is that some of these alliinase gene family members may be functionally inactive. Non-functional pseudogenes can represent a significant proportion of a gene family, e.g. three of the eight genes encoding lectins in the castor oil plant, have been shown to not produce transcripts (Tregear & Roberts 1992). Hybridization of the Alli6 cDNA probe to multiple alliinase genes by Southern analysis indicates that this probe could be used to isolate the corresponding genomic clones from an *A. cepa* genomic library. This would eventually enable expression analysis of the individual gene family members and identification of possible pseudogenes.

4.3. Expression of alliinase in developing *A. cepa* seedlings

Little work has been carried out on the expression of alliinase in different tissues or developmental stages. The lack of flavour release from certain *A. cepa* tissues, predominantly the cortical pith tissues of the leaf blade and scape (Lancaster & Boland 1990) provides indirect evidence of variation in flavour precursors and/or alliinase expression in different *A. cepa* tissues. Freeman (1979) reported that alliinase activity in *A. cepa* (cv. Rijnsburger) seeds was less than 2% of that in bulbs, and increased rapidly during seedling development to reach a stable maximum, 15 to 20 days after germination. In *Acacia georginae* and *Albizzia julibrissin*, higher levels of alliinase activity were detected in the hypocotyl than in the cotyledon or shoot of seedlings (Mazelis & Fowden 1973). For these reasons, northern analyses were performed to determine whether the reported changes in alliinase activity during seedling development could be correlated to changes in alliinase transcript abundance.

Northern blot analysis employing the Alli6 cDNA probe established that alliinase genes were transcriptionally active in *A. cepa* seedlings. Strong autoradiographic signals were detected from 10 μg of total seedling RNA even after four hour exposures (see Subsection 3.8 and Figure 3.16, page 146), indicating that alliinase transcripts were highly abundant at all the developmental stages examined. *A. cepa* alliinase is believed to be a multimer composed of identical 50 kDa subunits (Tobkin & Mazelis 1979, Nock & Mazelis 1987). This is consistent with the single transcript of about 1.7 kb detected in all the RNA samples probed. The alliinase cDNA clones ranged in length from 1.60 kb (Alli4B) to 1.76 kb (Alli7) (see Figure 3.8), indicating that these were similar in length to intact alliinase transcripts.

Spectrophotometric estimates of RNA concentration appeared unequal as judged by visual comparison of samples on ethidium bromide-stained agarose/formaldehyde gels. This suggested that RNA prepared from *A. cepa* seedlings may contain a contaminant which affected absorbance at 260 nm. The co-purification of unidentified compounds which absorbed light at wavelengths similar to those of nucleic acids has also been described in RNA extractions from grape berry tissues (Tesniere & Vayda 1991). To correct for possible RNA loading variation and for variations in the efficiency of RNA transfer to membranes, the intensity of Alli6 binding to each sample was calculated relative to that of a *Malus* 45S rDNA probe (Simon & Weeden 1992). Ribosomal RNA sequences are extremely well conserved

across taxa, hence *A. cepa* and *Malus* rRNA sequences would be expected to be highly homologous. Restriction enzyme analysis shows the organization of multiple restriction enzyme sites within the transcribed regions of the nuclear 45S rDNAs of six *Allium* species (Havey 1992) and two *Malus* genotypes (Simon & Weeden 1992) are identical. Ribosomal RNAs represent a constant fraction of mammalian RNA (Leeuw *et al.* 1989), whereas the expression of 'housekeeping' genes can be subject to biological variation. rRNAs are therefore more reliable as standards for RNA loading and transfer (Correa-Rotter *et al.* 1992). Consequently, correction of alliinase probe hybridization intensities relative to rDNA binding should give values more accurately reflecting transcript abundances in the total RNA population.

The relative abundances of alliinase transcripts at five stages of *A. cepa* seedling development were found to differ (see Subsection 3.8 and Figure 3.16), indicating that alliinase expression may be transcriptionally regulated. Transcript abundance was highest during root and cotyledon emergence (Stages 1 and 2), and remained high through to the stage of adventitious root development (Stage 3). By the time the first true leaf emerged (Stage 5) the proportion of alliinase transcripts had declined substantially, suggesting a drop in the rate of alliinase gene transcription during seedling development. Therefore, alliinase transcripts appear to be most abundant at the time of emergence of the root, hypocotyl and early cotyledon tissues of developing *A. cepa* seedlings. The increase in transcript abundance from Stage 1 to Stage 2 implies that alliinase transcripts may be present at lower levels in seeds and that alliinase genes become transcriptionally active during or after germination. However, analysis of alliinase transcript levels in seeds is required to confirm this. The pattern observed was independent of light, being similar in both the light and dark grown seedlings, but the slightly lower overall abundance of alliinase transcripts in dark-grown tissue (Figure 3.17) suggests light may exert some independent effect on transcription of alliinase gene(s).

Under the conditions used, the alliinase transcripts detected by northern analysis may have arisen from any of the several homologous genes inferred by Southern blot analysis of *A. cepa* DNA (discussed in the previous section). Consequently it is not possible to discern from these data whether the observed changes in alliinase transcript abundance are due to variation in the transcription of a single alliinase gene, or to the combined expression patterns of multiple gene family members. Characterization of more cDNA and genomic clones is required before gene-specific probes, which will allow a detailed examination of alliinase transcription, can be designed. A more thorough analysis of the expression of alliinase may contribute

to an understanding of this enzyme's biological function.

The pattern of alliinase accumulation reported in *A. cepa* seedlings by Freeman (1979) together with the pattern of alliinase transcript abundance detected in this study would suggest alliinase protein is turned over at a slow rate, as activity levels were reported to remain stable well after transcript abundance was found to decline. However ,the possibility that the abundance of alliinase transcripts may increase in tissues older than 14-15 days cannot be dismissed. That *A. cepa* alliinase is highly stable has also been implied by the results of protein turnover studies using radiolabelled amino acids (U. Wäfler pers. comm.). Low protein turnover is consistent with the observation that alliinase is highly abundant in dormant bulb tissue (Nock & Mazelis 1987), in which transcription and protein synthesis rates would be predicted to be low. Further studies are required to elucidate the precise relationship between alliinase transcript abundance and alliinase activity.

Alliinase activity in seedlings of *Acacia georginae* and *Albizzia julibrissin* has been reported to be higher in the hypocotyl and root of seedlings than in the cotyledon (Mazelis & Fowden 1973), which compares favourably with the timing of the changes found here. An abundance of alliinase and its substrates in seedlings could result in the release of significant levels of antimicrobial thiosulphinates (Virtanen 1965, Fenwick & Hanley 1986) when these tissues are wounded. The pattern of alliinase transcript accumulation found in this study also correlates with the pattern of *S*-alk(en)yl-CSO flavour precursor accumulation described in *A. cepa* seedlings by McCallion & Lancaster (1984). CSOs were not detectable in *A. cepa* seeds, but were found to accumulate during the four days after germination and then declined rapidly during the ensuing ten days (McCallion & Lancaster 1984). It has been suggested that an accumulation of alliinase during germination and seedling development may function in the successful establishment of seedlings by providing a means of defence against soil borne pathogens (Mazelis & Fowden 1973).

Chapter 5. Concluding summary and discussion

5.1. Summary of clone isolation and identification

The principle aim of this project was to isolate an alliinase clone. Alliinase-mediated hydrolytic cleavage of the *S*-alk(en)yl-CSO precursors is central to the release of flavour by *Allium* species. It was envisaged that a clone encoding alliinase would provide a molecular tool that would facilitate investigations into the expression, activity and function of this novel vacuolar enzyme, and in the longer term, allow the development of strategies aimed at altering onion flavour. To achieve this end, amino acid sequences from purified alliinase were obtained. It was anticipated that back-translation of these peptides would allow the production of oligonucleotides that could be used to isolate an alliinase clone by homology probing of a gene library or by PCR amplification. In addition, alliinase amino acid sequences would also be valuable in confirming clone identity. Furthermore, anti-alliinase antisera were raised against the purified native enzyme with the intention of screening for immunologically reactive proteins expressed by clones within an *A. cepa* cDNA library.

5.1.1. Summary of the use of redundant oligonucleotides for clone isolation by homology probing and PCR

Two oligonucleotides (JL1 and JL2) of limited redundancy were synthesized for library screening after back-translation of a 23-residue N-terminal *A. cepa* alliinase peptide sequence (ALN1, see Table 3.1). Codon choices were made by considering the preference of monocot genes for G or C in the third base position, as described by Murray *et al.* (1989) (described in Section 4.1.1 and presented in Figure 2.1). Both hybridization of an *A. cepa* genomic library constructed in λEMBL3 with these oligonucleotides, and subsequent screening with a putative alliinase cDNA probe (described in Sections 3.2.2 and 3.6) failed to identify any potential alliinase clones.

It was considered that the small size of the gene library and the high redundancy of the probe sequences (see discussion in Section 4.1.1) were either individually or in combination responsible for this negative outcome. A possible cause for the insufficient size of the gene library was later found to be the use of the *E. coli* host strain, LE392, for library screening. At least two loci of *E. coli* LE392 encode nucleases that specifically recognize DNA containing methylated cytosine in certain

sequence contexts. This strain is known to substantially lower the titre of genomic libraries constructed with plant DNAs, which generally have high levels of m^5C (Graham *et al.* 1990). It was calculated that the genomic clone bank constructed in this study represented approximately 5% of the entire *A. cepa* genome (discussed in Subsection 4.1.1). As DNA sequences encoding alliinase were subsequently found to be represented by probably less than ten copies per *A. cepa* genome (described in Section 3.7), it was considered that the small gene bank created may not have contained an alliinase clone.

The codon redundancies of the oligonucleotide probes, although approximating those used successfully for clone isolation by library screening in some previous studies (see discussion in Subsection 4.1.1), were initially considered to be too great to provide a readily detectable autoradiographic signal given the specific activities achieved. Subsequent isolation and DNA sequencing of alliinase cDNA clones allowed an evaluation of the sequence homology of these probes. It was found that the reliance on the Murray *et al.* (1989) survey of monocot codon usage to make codon choices resulted in a lack of absolutely complementary probes among the redundant sequence options of the JL1 and JL2 oligonucleotides (see discussion in Subsection 4.1.1). The most complementary of these oligonucleotide sequences contained two base mismatches, and this may account for the difficulties encountered in achieving authentic stable probe-target hybrids. It is apparent that trends derived from the predominantly graminaceous G+C-rich monocot genes sequenced to date cannot be assumed to apply to the DNA sequences of taxonomically distant monocot species. Campbell & Gowri (1990) found that monocot genes fall into two classes, one with sequences containing a high proportion of G or C in third base position, and another with base compositions less G+C-rich, and more like those usually reported for dicot genes. The codon usage of the subsequently isolated alliinase cDNAs displayed a base composition more comparable with that of the latter group and also with the overall base composition of *A. cepa* genomic DNA reported by Kirk *et al.* (1970). It appears that genomic DNA composition data may be of value when attempting to limit the degeneracy of synthetic oligonucleotide probes designed from peptide sequences.

Eight further oligonucleotides, six of which encompassed all possible redundant sequence options, were designed from the N-terminal alliinase peptide sequence (ALN1), and the sequence of a cyanogen bromide-generated internal fragment (ALN3) whose location within the protein was unknown (Figure 2.1). These oligonucleotides were used to prime PCR amplifications from *A. cepa* cDNA and genomic DNA templates using several strategies such as nested primers and template

selection by Southern hybridization, which were intended to enhance specificity (described in Section 3.3). No alliinase-encoding sequences were isolated (discussed in Section 4.1.2).

Initial selection of potential authentic fragments on the basis of predicted size led to the isolation of a 1.5 kb product that was amplified from cDNA templates with the sense primer JL5 and the non-specific primer $(dT)_{17}$. This was ligated into the plasmid pUC19 and partially DNA sequenced across both vector-insert borders. Both termini of this amplification product contained the sequence of the sense primer used in these reactions (JL5). The non-specific primer $(dT)_{17}$ was, however, essential for its amplification, suggesting that artifactual products may have been generated during PCR by product concatamerization (discussed in Section 3.3.1.1). Translation of these DNA sequences in the reading frame contiguous with that of the primer sequence clearly demonstrated that this product did not encode alliinase (presented in Figure 3.2, page 95). Numerous preliminary PCR studies were performed using a range of primers, templates and reaction conditions (described in Section 3.3) and generally resulted in non-specific amplification products. It was therefore concluded that the specificity of the degenerate oligonucleotides designed by back-translation of the alliinase peptide sequences available was insufficient to preferentially amplify alliinase-encoding sequences. PCR approaches were therefore discontinued. Later comparison of these primer sequences to the corresponding regions of an alliinase cDNA clone revealed several differences in their nucleotide sequences. These could be accounted for by amino acid sequence differences in the directly determined and cDNA-inferred alliinase proteins. It was suggested that this variation may have been due to the use of different cultivars for the isolation of pure alliinase and of mRNA for cDNA library construction (discussed in Section 4.1.2).

5.1.2. Summary of clone isolation by expression library immunoscreening

An *A. cepa* cDNA expression library was constructed in λZAPII and immunoscreened with polyclonal rabbit antisera raised against purified native alliinase. Western blotting of TFMS-treated alliinase was employed to demonstrate the affinity of the polyclonal antisera to the peptide back bone of this glycoenzyme (described in Subsection 3.4.2). The possibility that anti-alliinase polyclonal sera may contain antibodies directed against other plant (non-alliinase) epitopes was never ignored. DNA sequencing of immunoreactive clones later revealed that the polypeptide expressed by the clone pAlli4A showed no similarity to the alliinase peptide sequences (discussed in Subsection 4.1.3.1). In total, seven immunopositive clones were identified and excised *in vivo* into the phagemid pBluescript(-) (described

in Subsection 3.4.3). Initial characterization of the relationship of these cDNAs was by digestion mapping with the restriction endonucleases *Eco* RI and *Xho* I (described in Section 3.5.1). Resultant fragment sizes calculated from electrophoretic migration on agarose minigels indicated that six of the seven cDNAs contained an *Eco* RI site approximately 200 bp from the 5' end (presented in Figure 3.5). Five of these clones appeared to be between 1.6 and 1.7 kbp, and one, Alli9, was 600-700 bp longer. The seventh clone, Alli4A (which was later shown to be unrelated), was approximately 0.8 kb in length and did not contain any internal *Xho* I or *Eco* RI sites. As the restriction maps of five of these clones appeared to be very similar, further analysis was carried out by DNA sequencing. Double-stranded nucleotide sequencing of the cDNA inserts within the pBluescript phagemid vectors was initially primed with universal pUC/M13 forward and SK primers from the flanking vector sequences, and was continued with synthetic oligonucleotides designed from the cDNA sequence as it was ascertained.

Three clones (Alli4A, 4B and 6) were sequenced entirely, whereas the clones Alli7, 8B and 9 were incompletely sequenced from their termini (this strategy is illustrated in Figure 3.7). A further clone, pAlli8, was not DNA sequenced but appeared indistinguishable from pAlli4B on the basis of the restriction fragment sizes generated with the endonucleases *Eco* RI and *Xho* I (shown in Figure 3.5 on page 115). Four of the clones (Alli6, 7, 8B and 9) which had appeared to have restriction enzyme sites in similar locations, were found to contain identical DNA sequences (presented in Figure 3.8), with the only variation found being the length of the 5' ends. These sequence data confirmed the location of an *Eco* RI site near the 5' end, at position 328 of the Alli7 sequence (presented in Figure 3.9, page 123). The clone Alli4B was indistinguishable in DNA sequence from Alli6, except for a truncated 3' untranslated region with 6 different bases immediately preceding the polyadenylate tail. The differences found clearly demonstrated that these clones were all independently isolated. An additional sequence was determined from a cDNA which appeared to have become incorporated into the clone pAlli9 by artifactual ligation of two sequences during library construction. Comparison of this sequence to those filed within the GenBank database indicated a significant sequence resemblance to the cyclophilins of *B. napus* and *L. esculentum* (see Subsections 3.5.2 and 3.5.2.4).

The entire longest open reading frame of the five nearly identical cDNA sequences encoded a 479 residue protein. It was found that the sequences of all but one of the alliinase peptides presented in Table 3.1 aligned with the inferred polypeptide. The deduced protein sequence varied in only four residues from the 59 amino acids determined directly from the N-terminus of the native enzyme and three

cyanogen bromide-generated peptide fragments (presented in Figure 3.8). The peptide sequence ALN2, which could not be aligned with the sequences inferred from the cDNA clones, was found to show significant identity with the lectins of *A. sativum* (see Subsection 3.1), indicating that the alliinase preparation used for amino acid sequencing contained a co-purifying lectin. Furthermore, the cDNA-derived polypeptide sequence also displayed consensus sequence motifs for Asn-linked glycosylation and PLP binding which are consistent with these clones encoding alliinase (described in Subsections 3.5.2.2 and 3.5.2.3). These data provide strong evidence that five of the six clones isolated and sequenced in this study encode an *A. cepa* alliinase subunit precursor. No significant homologies were found between the sequences of these alliinase clones and any entry within the EMBL or GenBank databases.

The alliinase subunit precursor encoded by these clones was predicted to have a molecular mass of 54.8 kDa and contained 34 residues at the N-terminus prior to the site of alignment of the N-terminal sequence of mature alliinase. This preprotein displayed characteristics consistent with those of an ER targeting sequence, in that it was notably hydrophobic and was separated from the mature alliinase protein by a signal peptidase recognition site (discussed in Subsection 4.1.3.3.1). Vacuolar proteins such as alliinase are believed to be secreted cotranslationally into the ER, and subsequently translocated via the secretory system to their final destination (Boller & Wiemken 1986). Removal of this targeting signal sequence results in a predicted alliinase protein of 51.4 kDa, somewhat larger than the 47 kDa estimated for deglycosylated *A. cepa* alliinase by SDS-PAGE (discussed in Subsection 4.1.3.3.2). Analysis of the proteins expressed from the Alli6 cDNA in *E. coli* by western blotting resulted in the detection of a major band of 47 kDa, and a further immunoreactive protein of 41.4 kDa. The smaller polypeptide could be accounted for by translational initiation from an internal methionine residue at position 124. The 47 kDa band could not, however, be similarly explained (presented in Figure 3.12). Whether this discrepancy in the predicted and expressed polypeptide sizes is the result of post-translational modification, proteolytic cleavage, anomalous migration or fragmentation during SDS-PAGE, could not be resolved.

The Alli7 cDNA contained a 111 bp 5' leader region that was not found within the DNA sequences of the four other alliinase clones. Multiple translation start and stop codons resulted in the presence of several small open reading frames within this leader sequence (presented in Figure 4.1), suggesting that some form of ribosomal reinitiation might be required for translation of this transcript to proceed *in vivo* (see discussion in Subsection 4.1.3.2.2). Upstream open reading frames in eukaryotic

transcripts are rare (Kozak 1986) and have been described in only a few plant mRNAs, such as the *phyA* gene transcripts of several dicotyledonous plants (Sharrock & Quail 1989). The unusual structure of this region, together with its absence from the other independently isolated alliinase cDNAs, suggests that its presence within the Alli7 clone may have been generated artifactually during library construction. The cDNA clones described in this work have since been used to isolate alliinase clones from an *A. cepa* genomic library by B. Gilpin (PhD student). DNA sequence data derived from one alliinase genomic clone has demonstrated that no sequences homologous to the Alli7 cDNA are found immediately 5' to position 112 (B. Gilpin pers. comm.), consistent with the hypothesis that the Alli7 leader sequence may be artifactual. Further DNA sequence data from 5' regions of other *A. cepa* alliinase genomic clones, or 5'-end mapping studies of alliinase transcripts will be required to clarify this point.

5.2. Analysis of the genomic organization and expression of alliinase genes in *A. cepa*

Southern analysis of *A. cepa* DNA digested with seven restriction endonucleases was carried out with the cDNA clone Alli6 (presented in Figures 3.13 and 3.14). An accurate estimate of the alliinase gene copy number was complicated by variations in the number of hybridizing bands and probe binding intensities. Nevertheless it was proposed that alliinase is probably encoded by a small gene family of at least four members (see discussion in Subsection 4.2.2.1). Isolation and DNA sequencing of the corresponding genomic sequences are necessary to unequivocally confirm this finding. Two genomic clones have been recently isolated by screening an *A. cepa* genomic library in λGem12 with the cDNA probe Alli6. These appear to contain introns of different sizes, as well as divergent 3' non-transcribed sequences, supporting the hypothesis that multiple loci encode alliinase in *A. cepa* (B. Gilpin pers. comm.). Southern analysis with a 3' untranslated region alliinase cDNA probe, Alli6-3' (presented in Figure 3.13) suggested that the cDNA clones isolated in this study all originate from either a single gene family member, or two very closely related, possibly tightly linked, genes, which is consistent with their considerable sequence similarity. The cDNA expression library from which these clones were isolated was constructed with mRNA extracted from sprouting shoots of *A. cepa* bulbs (see Subsection 2.4.4.1.1). This, together with the DNA hybridization data discussed above, suggests that either some alliinase gene family

members are functionally inactive, or that their transcription is regulated in a tissue-specific or developmentally regulated fashion (see discussion in Subsection 4.2.2.3).

This hypothesis was investigated by using the clone Alli6 to probe northern blotted light- and dark-grown *A. cepa* seedling RNA from five stages of early development which spanned the period from germination to the emergence of the first true leaf (presented in Figure 3.16). Alliinase transcripts in both light- and dark-grown tissues were found to become highly abundant within the 2-3 days following germination and remain high for 5 to 6 days. However, transcript levels rapidly declined over the ensuing 4-8 day period. These data suggest that alliinase expression is probably regulated at the level of transcription (see discussion in Section 4.3). This result is also in agreement with the study of Freeman (1979) in which alliinase activity was not detectable at significant levels in *A. cepa* seeds, but rapidly increased immediately following germination. It was not possible, however, to determine if the alliinase mRNA abundance observed was due to transcription of a single or multiple alliinase gene(s). Moreover the mechanism of induction of alliinase transcription is as yet unknown. It is anticipated that DNA sequence analysis of genomic clones will enable the production of probes specific to individual gene family members. The application of such probes in RNA hybridization experiments or ribonuclease protection assays will allow a more detailed examination of alliinase gene expression patterns and may provide clues as to the signals important in the induction of alliinase transcription.

5.3. Future prospects

No genes encoding alliinase or closely related enzymes have been previously isolated. Nock & Mazelis (1986, 1987, 1989) have reported physicochemical, kinetic and immunological data that suggest differences exist in the structure and active sites of *A. cepa* and *A. sativum* alliinases. A comparison of the sequences of alliinase and alliinase-like enzymes from a variety of *Allium* and non-*Allium* species would permit an analysis of protein sequence differences and, it would be hoped, ultimately lead to the identification of biologically important domains. The four amino acids at the N-terminus of native *A. sativum* alliinase have been identified, and these show a strong resemblance to the corresponding sequence of the *A. cepa* enzyme (Nock & Mazelis 1989) implying further sequence homologies may also exist. This study has therefore provided a cDNA probe which could allow the isolation of alliinase gene from a range of species, and will ultimately enable detailed analyses to be carried out on the structure, function and regulation of alliinase.

5.3.1. Expression of cDNA clones in *E. coli* or *in vitro*

The evidence that the five cDNAs Alli4B, 6, 7, 8B and 9 encode alliinase relies primarily on the identification of the alliinase peptide sequences and the specificity of the anti-alliinase antisera. That the clones isolated encode alliinase assumes that the protein used both for antibody production and peptide sequencing was in fact alliinase and not a copurifying contaminant. Support for their identity is the agreement in the characteristics of the cDNA-inferred protein sequence and those of purified native *A. cepa* alliinase. In addition, the authenticity of the peptide sequences is indicated by the similarity of the four N-terminal residues to those previously determined from *A. sativum* alliinase by Nock & Mazelis (1989) (see Subsection 3.1), by the high specific activity of the alliinase preparation used in this study, and the presence of a single protein band after SDS-PAGE and silver staining. Western blot analysis of deglycosylated alliinase demonstrated the affinity of the anti-alliinase antisera to epitopes present within the alliinase protein (Subsection 3.4.2). However, the isolation of an unrelated clone, Alli4A, indicated that the anti-alliinase polyclonal preparation could recognize epitopes presented by other *A. cepa* proteins (discussed in Subsection 4.1.3.1).

Further experiments that would unequivocally confirm the identity of the clones isolated in this study could be based on the expression of a full-length cDNA in a heterologous system. Detection of alliinase activity would therefore be directly attributable to the presence of the cloned sequence. Heterologous expression of a clone at high levels would also provide a means to obtaining large amounts of pure alliinase protein for functional studies. Both *A. cepa* and *A. sativum* alliinases are believed to be homomeric as evidenced by their migration on denaturing PAGE as single bands (Tobkin & Mazelis 1979, Nock & Mazelis 1987) and the absence of significantly different isoforms (Nock & Mazelis 1987, Jansen *et al.* 1989a). This is also supported by the isolation in this study of five independent clones that all encode an identical alliinase subunit polypeptide to which each of the known alliinase peptide sequences could be aligned. The alliinase cDNA clones all appear to encode the entire subunit precursor protein, consequently the protein expressed from a single cDNA may be capable of assembling to form an intact, functional, alliinase holoenzyme.

Several cDNA clones of plant enzymes have been introduced into *E. coli* and shown to express catalytically active forms. Udvardi & Kahn (1991) detected the activity of a *Medicago sativa* PLP-requiring enzyme, aspartate aminotransferase (AAT), when a cDNA was cloned into the pUC18 expression vector and expressed in *E. coli*. Activity was directly detected within aqueous extracts of the transformed cells by an AAT-malate dehydrogenase coupled assay system which allows the reaction to be followed via the spectrophotometric determination of NADH

oxidation. The assembly of active plant enzymes such as glutamine synthase (Bennett & Cullimore 1990) and parsley phenylalanine ammonia lyase (Schulz *et al.* 1989) in *E. coli* from cDNA clones demonstrates that successful formation of the holoenzymes can occur without the presence of plant-specific assembly factors or multiple gene copies. A further approach for *in vitro* expression of catalytically active plant enzymes is the method of Reddy *et al.* (1987) in which a full-length cDNA clone encoding dihydroquercitin reductase, the A1 locus of *Zea mays*, was transcribed *in vitro* and translated in the wheat germ or reticulocyte lysate systems. Reductase activity was demonstrated and thus confirmed clone identity in a functional manner.

These and other examples of catalytically active plant enzyme expression in heterologous systems require no post-translational processing other than that functioning in *E. coli* or the *in vitro* translation systems. This does not include modifications such as glycosylation. The covalent attachment of oligosaccharide side chains to selective asparagine residues is carried out within the lumen of the rough ER (Kornfeld & Kornfeld 1985). Although the function of glycans in plant glycoproteins has not been fully elucidated, it is believed they may have a role in stabilizing protein conformation or providing protection from proteolytic degradation (Faye & Chrispeels 1988) and may therefore be essential for biological activity. The deglycosylation by 40% of an anionic isoperoxidase with glycopeptidase F, for instance, has been shown to result in the loss of up to 78% of the enzyme activity, as well as a significant reduction in enzyme stability (Tigier *et al.* 1991). A *Phaseolus vulgaris* gene encoding a glycosylated α-amylase inhibitor was found to be non-functional when cloned in frame with the *lacZ* gene of a pBluescript(-) phagemid and expressed in *E. coli* (Moreno & Chrispeels 1989). It was thought that appropriate proteolytic processing and/or glycosylation were essential for the acquisition of activity. The expression of an active alliinase holoenzyme from the cDNA clones isolated in this study and resultant confirmation of their identity by functional assay, may therefore require appropriate glycosylation within an intact eukaryotic expression system.

5.3.2. Expression of alliinase in heterologous plant systems

The development of efficient systems for plant cell transformation has provided an experimental methodology by which gene constructs can be evaluated in transformed protoplasts, calli or intact plants. Alliinase and alliinase-like enzymes

appear to be limited to relatively few plant families such as the *Alliaceae* (Lancaster & Boland 1990), *Amaryllidaceae* (Jacobsen *et al.* 1968), *Brassicaceae* (Mazelis 1963) and *Fabaceae* (Mazelis & Creveling 1975). Hence many species commonly in present use as plant transformation systems are unlikely to show endogenous alliinase, or alliinase-like, activity. Confirmation of this could be easily obtained by direct alliinase activity assay by the method of Schwimmer & Mazelis (1963). Therefore the use of plant transformation technology could provide unequivocal confirmation of the identity of these cDNAs, as expression of alliinase by transgenic tissues would be due solely to the introduction of the cloned sequence. Translation products expressed by genes isolated from divergent sources are often correctly modified and compartmentalised when introduced into foreign plant tissues. For example a chimeric gene encoding the N-terminal portions of the potato proteinase inhibitor II and the coding sequence of the yeast invertase gene *suc 2* were fused to a 35S CaMV promoter and expressed in tobacco and *Arabidopsis thaliana* using *Agrobacterium*-based transformation techniques. Invertase was correctly glycosylated at Asn-X_{aa}-Ser/Thr sites and efficiently secreted within regenerated transgenic tobacco tissues, which displayed up to 500-fold higher invertase activity than non-transformed control plants (von Schaewen *et al.* 1990).

Plant transformation experiments could also provide a means of identifying the determinants active in targeting alliinase to the vacuole. Wilkins *et al.* (1990) have demonstrated appropriate targeting of barley lectin to the vacuoles of transgenic tobacco, and the bean seed lectin phytohemagglutinin is also correctly localised in the vacuole when expressed in yeast (Tague & Chrispeels 1987), indicating a wide conservation exists in the function and recognition of vacuolar sorting elements among many eukaryotes. A general consensus among the sequences known to be active in targeting of plant vacuolar proteins is that they tend to contain a number of residues with hydroxyl groups (serine and threonine residues) interspersed with both negatively and positively charged residues, and that they interact in a positive manner (Holwerda *et al.* 1992). However elucidation of the essential conserved features has been complicated by the diversity of peptide sequences exhibited by the vacuolar targeting determinants so far identified. Transformation experiments with constructs containing deletion mutants of the alliinase coding region, and subsequent localization of the expressed proteins by subcellular fractionation experiments, would enable regions essential for vacuolar compartmentalization of alliinase to be identified.

5.3.3. Antisense strategies

The ability to switch off or inhibit specific gene expression in plant cells has enormous potential. Plant genes can be inhibited or downregulated by anti-sense RNA and the use of this technology could permit the identity of the putative alliinase clones to be unequivocally confirmed. This technique would allow the isolated genes to be assigned a particular function on the basis of altered phenotype. A construct containing the gene of interest in reverse orientation is placed under the control of a promoter such that an antisense RNA complementary to the endogenous mRNA is transcribed. A double stranded hybrid is thought to be formed, which either prevents mRNA transport into the cytoplasm, inhibits translation or promotes degradation of the RNA-mRNA hybrid by ribonuclease action (Rothstein *et al.* 1987).

The first example of anti-sense RNA mediation of plant gene expression was reported in 1987 when Rothstein *et al.* (1987) placed the 5' half of the nopaline synthase (NOS) coding region in an anti-sense orientation under the control of the CaMV 35S promoter and introduced this construct into a NOS-expressing tobacco line using *Agrobacterium*-mediated transformation methods. An 8-50 fold decrease in NOS activity was demonstrated. Since then, antisense approaches have resulted in successful downregulation of various plant enzymes such as polygalacturonase in tomatoes (Smith *et al.* 1988) and chalcone synthase in *Petunia hybrida* (van der Krol *et al.* 1988) among others. Antisense strategies provide powerful tools to analyse the function of endogenous enzymes in cellular metabolism and plant development. For example, Neuhaus *et al.* (1992) demonstrated a 20-fold reduction in the constitutive expression of vacuolar class I β-1,3-glucanase antigen in tobacco leaves following the introduction of an antisense construct, but found no change in the susceptibility of these plants to infection by the fungal leaf pathogen *Cercospora nicotianae*. Antisense downregulation of chalcone synthase expression in anther tissue and pollen of petunia was shown to inhibit flavonoid synthesis which led to arrested pollen maturation and the production of male sterile plants, demonstrating that flavonoids play an essential role in male gametophyte development (van der Meer *et al.* 1992b).

For effective regulation by antisense RNA, the RNA-RNA interaction must occur during gene transcription (Eguchi *et al.* 1991). There are some indications that alliinase is a highly stable protein, exhibiting a relatively low rate of turnover (U. Wäfler pers. comm.). Alliinase is highly abundant in dormant tissues, such as *A. cepa* bulbs and *A. sativum* cloves, in which transcription may be predicted to be low.

Hence the alliinase activity levels detected in many *A. cepa* tissues may be due to a high abundance of accumulated protein rather than significant rates of transcription. To ensure that antisense constructs will be optimally expressed both spatially and temporally to maximise any biologically significant effect, it will be desirable to first establish the developmental and spatial patterns of transcription that result in alliinase production and accumulation throughout the plants life cycle. The results of the present study indicate that alliinase transcription varies in a developmental or tissue-specific way during early seedling development, which implies that transcriptional regulation may also play an important role in alliinase accumulation at other developmental stages. Alliinase transcripts appeared to be highly abundant during seedling development, and the use of a powerful constitutive promoter to drive the expression of an antisense construct may be required to provide sufficient complementary transcripts to block expression at all developmental stages. The high sequence homology of the alliinase genes detected by Southern analysis suggests that a single appropriately designed antisense sequence may be sufficient to bind to all alliinase transcripts expressed in *A. cepa*. If 5' and 3' gene-specific sequences were used then differential downregulation might be observed.

Although the examples discussed above demonstrate the potential of anti-sense RNA to alter the plant phenotype, considerable effort must be placed into developing the plant transformation technology before the experiments could be initiated. For this approach to be successful in the case of alliinase a transformation system in *A. cepa* would be preferred. As alliinase mediates the release of flavour in *Allium* species, it is anticipated that the development of an *A. cepa* transformation system will allow the use of strategies such as those discussed above to modify flavour release by altering alliinase expression. Development of transformation systems for monocot plants has proved to be difficult, in part due to the lack of susceptibility of many monocot plants to *Agrobacterium* infection and subsequent integration of the T-DNA (Dommisse *et al.* 1990). However, significant progress has been made in the application of other gene transfer techniques to previously recalcitrant monocot crop plants such as rice (Gasser & Fraley 1989). *A. cepa* has been demonstrated to act as a host for *Agrobacterium*, as evidenced by opine-positive tumourigenic responses to *Agrobacterium* inoculation of basal meristems (Dommisse *et al.* 1990), suggesting that current T-DNA transfer methodologies have the potential to facilitate transformation of *Allium* species. Furthermore, *Agrobacterium*-mediated transformation techniques have been shown by histochemical assay, to result in the expression of transgenes encoding β-glucuronidase (GUS) within *A. cepa* tissues, and to confer antibiotic resistance to the kanamycin analogue, G418 (Dommisse 1993).

Integration of transgenes has not yet been unequivocally demonstrated by molecular methods however, and it is believed that *Agrobacterium*-based methodologies employed to date may favour the production of chimeric *A. cepa* tissues. Other approaches to transformation such as microprojectile-mediated gene insertion (Klein *et al.* 1987) may be required to improve the quality of *A. cepa* transformation. Such techniques show considerable potential in promoting stable integration of transgenes in monocot species. Microprojectile bombardment has, for example, recently been shown to successfully transfer a gene for phosphinothricine resistance to embryogenic callus of the monocot *Avena sativum* (oat) (Somers *et al.* 1992).

5.3.4. Alliinase promoter analysis

A major limitation in the development of gene transfer techniques for monocot crop plants has been the lack of functionally optimal promoters. Although the CaMV 35S promoter is widely used in dicot plant transformation systems, it has been found to have relatively low activity in certain agronomically important monocot crop plants such as rice (McElroy *et al.* 1990c) and wheat (Hauptmann *et al.* 1987). Promoters isolated from the maize *adh1* gene (Last *et al.* 1991) and the rice *act1* gene (McElroy *et al.* 1990c) have been shown to give higher levels of expression in graminaceous monocots. These observations suggest that there may be differences between monocots and dicots with respect to transcription factors and the recognition of promoter sequences.

Alliinase is highly abundant in *A. cepa* and *A. sativum* tissues, representing up to 6% (w/w) of the soluble protein in *A. cepa* bulbs, and 12% (w/w) of the soluble protein in *A. sativum* cloves (Nock & Mazelis 1987). It was therefore believed that the transcription of the alliinase gene may be driven by a powerful promoter. Expression studies performed in this work suggest that alliinase transcripts were present in high abundance during early seedling development but that alliinase genes may be transcriptionally regulated in a developmental or tissue specific manner, and hence transcript abundance may be low at other stages or in other tissues. Vectors constructed with promoters derived from alliinase genes may therefore provide a transformation system capable of directing specific temporal or spatial regulation. Further studies are required to ascertain the expression pattern of alliinase throughout the plants life cycle, and alliinase genes may yet prove to contain *cis*-acting elements that can confer high level expression when placed in an appropriate sequence context. Work to date indicates that further attempts to integrate transgenes into the *A. cepa* genome would benefit from the use of a promoter which has optimal activity

in *A. cepa* (E. Dommisse pers. comm.). Hence alliinase promoter regions may be of particular value in developing *A. cepa* transformation systems and, in the longer term, in driving the transcription of genes introduced into *Allium* species to express desired traits.

Isolation of genomic clones would allow an analysis of promoter regions to identify elements involved in the regulation of alliinase transcription. Although this was an initial aim, this was not achieved by the present study. Comparison of 5' flanking sequences with consensus motifs known to act as regulatory promoter elements may locate and identify functional elements and imply environmental or developmental cues involved in transcriptional control.

Transient expression assays of chimeric constructs containing the intact or partially deleted 5' flanking sequences of alliinase genomic clones fused to a suitable reporter sequence such as GUS could be expressed in protoplasts to ascertain the functionality of various promoter elements. Furthermore, such constructs can be introduced into homologous or heterologous host genomes and their activity under various environmental and developmental conditions determined by histochemical localisation of GUS expression. These strategies have been used to identify several *cis*-acting regulatory elements active in plant gene expression. For example, a 38 bp poly(dA-dT) region within the *act1* gene promoter was identified as a positive regulator of *act1* activity by assaying the effect of this element on the expression of a downstream GUS gene in both rice protoplasts and transformed maize tissues (Wang *et al.* 1992). Deletion of this poly(dA-dT) element lowered GUS expression by at least three-fold compared with expression produced by the full-length *act1* promoter. Schmid *et al.* (1990) demonstrated that a bean CHS promoter conferred both developmentally and environmentally regulated expression patterns on a GUS gene introduced into transgenic tobacco. This promoter was highly active in greening cotyledons, primary leaves and in the root apical meristem and petals, as well as responding to wounding caused during lateral root formation, and to fungal elicitor treatment.

It is anticipated that future elucidation of the expression patterns conferred by alliinase promoters may provide some clues as to the signals active in the induction of alliinase transcription. *In situ* hybridization of *A. cepa* tissues using alliinase clones could further augment an analysis of the expression patterns of alliinase genes. In the longer term, it is also hoped to explore the molecular mechanism by which alliinase promoter elements exert their effects.

Acknowledgements

I am very grateful to my supervisors, Mrs Jane E. Lancaster and Dr. David W. M. Leung, for their support and guidance throughout the course of this research.

I am indebted to Dr. Gail M. Timmerman for her generosity with her time, her knowledge and her good advice on many aspects of this thesis, and to Martin Shaw without whose superb protein purification skills this project would have been much more difficult. Many thanks also to Elvira Dommisse for numerous discussions, advice and morale boosting during this project, as well as for proof reading the final draft of this work.

Thanks to all the staff of the Plant Improvement Division of Crop and Food Research, Lincoln, for providing a wonderfully congenial and cooperative work environment. I would particularly like to thank Dr. Flerida Carino, Julie Farrant, Tonya Frew, Brent Gilpin, Sue Lamont, Carolyn Lister, Allison Miller, Meeghan Pither and Dr. Urs Wäfler for their readiness to discuss and debate many aspects of the work during its course, their generosity with their reagents, and their excellent company which has made the last three years so enjoyable. Thanks also to Steve Burgham, Landcare Research Ltd., Lincoln, and Dr Michael Boland, Dairy Research Institute, Palmerston North, for reading and commenting on Chapter 1, and Dr. Frank Sin for his helpful suggestions on the writing of this thesis.

I would like to acknowledge the aid of Robert Lamberts, Nigel Mahoney, Jonathon Poff, Sabrina Malcolm and Dougal Holmes for assistance with the production of the figures and diagrams, Margot Bowden for efficiently locating so many obscure articles, and Dr. Chris Frampton for advice on statistics.

Thanks to Richard Hyder and Dr. Randall Allardyce for preparation and titration of polyclonal sera, Drs. Allan Carne and David Christie for amino acid sequence determinations, and Dr Sue Gardiner for the gift of *Malus* 45S probe DNA.

Financial support for this work was supplied by a Crop and Food Research Ltd., NZ. Fellowship and a Crop Research Division Outstanding Achievement Award.

A special thanks to my family and friends for their support, and in particularly the love, support and patience of Tony Green is deeply appreciated.

References

Altschul S.F.A., Gish W., Miller W., Myers E.W. and Lipman D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

An G., Mitra A., Choi H.K., Costa M.A., An K., Thornburg R.W. and Ryan C.A. 1989. Functional analysis of the 3' control region of the potato wound-inducible proteinase inhibitor II gene. *The Plant Cell* 1:115-122.

Anderson M.L.M. and Young B.D. 1988. Quantitative filter hybridisation. In: Hames B.D. and Higgins S.J., editors. *Nucleic acid hybridisation - a practical approach*, pages 73-110. IRL Press, Oxford.

Anderson N.W. and Thompson J.F. 1979. Cystine lyase: β-Cystathionase from turnip roots. *Phytochem.* 18:1953-1958.

Angenon G., Van Montagu M. and Depicker A. 1990. Analysis of the stop codon context in plant nuclear genes. *FEBS Lett.* 271:144-146.

Antequera F. and Bird A.P. 1988. Unmethylated CpG islands associated with genes in higher plant DNA. *EMBO J.* 7:2295-2299.

Apuya N.R. and Zimmerman J.L. 1992. Heat shock gene expression is controlled primarily at the translational level in carrot cells and somatic embryos. *The Plant Cell* 4:657-665.

Aramuganathan K. and Earle E.D. 1991. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9:208-213.

Armstrong G.A., Weisshaar B. and Hahlbrock K. 1992. Homodimeric and heterodimeric leucine zipper proteins and nuclear factors from parsley recognize diverse promoter elements with ACGT cores. *The Plant Cell* 4:525-537.

Austin S.J. and Schwimmer S. 1971. L-γ-Glutamyl peptidase activity in sprouted onion. *Enzymologia* 40:273-285.

Baird W.V. and Meagher R.B. 1987. A complex gene superfamily encodes actin in petunia. *EMBO J.* 6:3223-3231.

Bairoch A. 1992. PROSITE: A dictionary of sites and patterns in proteins. *Nucleic Acids Res.* 20:2013-2018.

Beck B.N. and Ho S.N. 1988. Increased specificity of PCR-amplified products by size-fractionation of restriction enzyme-digested template genomic DNA. *Nucleic Acids Res.* 16:9051.

Becker A. and Schuphan W. 1975. Ein beitrag zur biogenese und biochemie antimikrobiell wirkender ätherischer öle der küchenzwiebel (*Allium cepa* L.). *Qual. Plant.* 25:107-169.

Bednarek S.Y. and Raikhel N.V. 1991. The barley lectin carboxyl-terminal propeptide is a vacuolar protein sorting determinant in plants. *The Plant Cell* 3:1195-1206.

Benfey P.N. and Chua N.-H. 1990. The cauliflower mosaic virus 35S promoter: Combinatorial regulation of transcription in plants. *Science* 250:959-966.

Bennett M. and Cullimore J. 1990. Expression of three plant glutamine synthetase cDNA in *Escherichia coli* - formation of catalytically active isoenzymes, and complementation of a *glnA* mutant. *Eur. J. Biochem.* 193:319-324.

Bernhard R.A. 1970. Chemotaxonomy: Distribution studies of sulfur compounds in *Allium*. *Phytochem.* 9:2019-2027.

Bernier G. 1988. The control of floral evocation and morphogenesis. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 39:175-219.

Berry J.O., Breiding D.E. and Klessig D.F. 1990. Light-mediated control of translational initiation of ribulose-1,5-biphosphate carboxylase in amaranth cotyledons. *The Plant Cell* 2:795-803.

Binnie C. 1991. Designing the optimal oligo. *TIBtech.* 9:145-146.

Bird C.R., Smith C.J.S., Ray J.A., Moureau P., Bevan M.W., Bird A.S., Hughes S., Morris P.C., Grierson D. and Schuch W. 1988. The tomato polygalacturonase gene and ripening-specific expression in transgenic plants. *Plant Mol. Biol.* 11:651-662.

Block A., Dangl J.L., Hahlbrock K. and Schulze-Lefert P. 1990. Functional borders, genetic fine structure and distance requirements of *cis* elements mediating light responsiveness of the parsley chalcone synthase promoter. *Proc. Natl. Acad. Sci. USA* 87:5387-5391.

Block E. 1985. The chemistry of garlic and onions. *Sci. Am.* 252:94-99.

Block E. 1992. The organosulfur chemistry of the genus *Allium* - implications for the organic chemistry of sulfur. *Angew. Chem. Int. Ed. Engl.* 31:1135-1178.

Boelens M., de Valois P.J., Wobben H.J. and van der Gen A. 1971. Volatile flavor compounds from onion. *J. Agr. Food Chem.* 19:984-991.

Boller T. and Wiemken A. 1986. Dynamics of vacuolar compartmentation. *Ann. Rev. Plant Physiol.* 37:137-164.

Boudraa M. and Perrin P. 1987. CpG and TpA frequencies in the plant system. *Nucleic Acids Res.* 15:5729-5737.

Breathnach R. and Chambon P. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50:349-383.

Brodnitz M.H. and Pascale J.V. 1971. Thiopropanol *S*-oxide: A lachrymatory factor in onions. *J. Agr. Food Chem.* 19:269-272.

Brodnitz M.H., Pascale J.V. and Van Derslice L. 1971. Flavor components of garlic extract. *J. Agr. Food Chem.* 19:273-275.

Brooks J.E. 1987. Properties and uses of restriction endonucleases. *Methods Enzymol.* 152:113-129.

Brown C.M., Stockwell P.A., Trotman C.N.A. and Tate W.P. 1990. Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. *Nucleic Acids Res.* 18:6339-6412.

Brown J.W.S. 1986. A catalogue of splice junction and putative branch point sequences from plant introns. *Nucleic Acids Res.* 14:9549-9559.

Bustos M.M., Begum D., Kalkan F.A., Battraw M.J. and Hall T.C. 1991. Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. *EMBO J.* 10:1469-1479.

Callis J., Fromm M. and Walbot V. 1987. Introns increase gene expression in cultured maize cells. *Genes Dev.* 1:1183-1200.

Campbell W.H. and Gowri G. 1990. Codon usage in higher plants, green algae, and cyanobacteria. *Plant Physiol.* 92:1-11.

Cavener D.R. and Ray S.C. 1991. Eukaryotic start and stop translation sites. *Nucleic Acids Res.* 19:3185-3192.

Ceci L.N., Curzio O.A. and Pomilio A.B. 1992. γ-Glutamyl transpeptidase/γ-glutamyl peptidase in sprouted *Allium sativum*. *Phytochem.* 31:441-444.

Chase M.W. and Palmer J.D. 1989. Chloroplast DNA systematics of liliod monocots: Resources, feasibility, and an example from the Orchidaceae. *Amer. J. Bot.* 76:1720-1730.

Chen W. and Struhl K. 1988. Saturation mutagenesis of a yeast *his3* "TATA element": Genetic evidence for a specific TATA-binding protein. *Proc. Natl. Acad. Sci. USA* 85:2691-2695.

Chen W. and Struhl K. 1989. Yeast upstream activator protein GCN4 can stimulate transcription when its binding site replaces the TATA element. *EMBO J.* 8:261-268.

Chen W., Tabor S. and Struhl K. 1987. Distinguishing between mechanisms of eukaryotic transcriptional activation with bacteriophage T7 RNA polymerase. *Cell* 50:1047-1055.

Chen Z.-L., Pan N.-S. and Beachy R.N. 1988. A DNA sequence element that confers seed-specific enhancement to a constitutive promoter. *EMBO J.* 7:297-302.

Chen Z.-L., Schuler M.A. and Beachy R.N. 1986. Functional analysis of regulatory elements in a plant embryo-specific gene. *Proc. Natl. Acad. Sci. USA* 83:8560-8564.

Chou P.Y. and Fasman G.D. 1974. Prediction of protein conformation. *Biochem.* 13:222-245.

Chou Q., Russell M., Birch D.E., Raymond J. and Bloch W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Res.* 20:1717-1723.

Chrispeels M.J. and Raikhel N.V. 1992. Short peptide domains target proteins to plant vacuoles. *Cell* 68:613-616.

Christensen A.H. and Quail P.H. 1989. Structure and expression of a maize phytochrome-encoding gene. *Gene* 85:381-390.

Christensen A.H., Sharrock R.A. and Quail P.H. 1992. Maize polyubiquitin genes: Structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Mol. Biol.* 18:675-689.

Clark S.A., Shaw M.L., Leung D.W.M. and Lancaster J.E. 1993. Isolation and characterization of a cDNA clone encoding alliinase - the enzyme responsible for flavour release in onion (*Allium cepa*). *Plant Mol. Biol.* Submitted.

Clegg J.C.S. 1982. Glycoprotein detection in nitrocellulose transfers of electrophoretically separated protein mixtures using concanavalin A and peroxidase: Application to Arenavirus and Flavivirus proteins. *Anal. Biochem.* 127:389-394.

Colman A., Byers M.J., Primrose S.B. and Lyons A. 1978. Rapid purification of plasmid DNAs by hydroxyapatite chromatography. *Eur. J. Biochem.* 91:303-310.

Connelly S. and Manley J.L. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* 2:440-452.

Correa-Rotter R., Mariash C.N. and Rosenberg M.E. 1992. Loading and transfer control for northern hybridization. *Biotechniques* 12:154-158.

Cronquist A. 1968. *The evolution and classification of flowering plants*, page 516. Houghton Miflin, Boston.

Currah L. and Ockendon D.J. 1978. Protandry and the sequence of flower opening in the onion (*Allium cepa* L.). *New Phytol.* 81:419-428.

D'Alessio J.M., Noon M.C., Ley III H.L. and Gerard G.F. 1987. One-tube double-stranded cDNA synthesis using cloned M-MLV reverse transcriptase. *BRL Focus* 9:1-4.

Dahlgren R. and Bremer K. 1985. Major clades of the angiosperms. *Cladistics* 1:349-368.

Dahlgren R.M.T., Clifford H.T. and Yeo P.F. 1985. *The families of the monocotyledons*, pages 193-199. Springer-Verlag, Berlin.

Davey M.R., Rech E.L. and Mulligan B.J. 1989. Direct DNA transfer to plant cells. *Plant Mol. Biol.* 13:273-285.

Davis B.D. and Tai P.-C. 1980. The mechanism of protein secretion across membranes. *Nature* 283:433-438.

Dean C., Favreau M., Bedbrook J. and Dunsmuir P. 1989a. Sequences 5' to translation start regulate expression of petunia *rbcS* genes. *The Plant Cell* 1:209-215.

Dean C., Pichersky E. and Dunsmuir P. 1989b. Structure, evolution and regulation of *rbcS* genes in higher plants. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 40:415-439.

Dean C., Tamaki S., Dunsmuir P., Favreau M., Katayama C., Dooner H. and Bedbrook J. 1986. mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. *Nucleic Acids Res.* 14:2229-2240.

Dean C., van den Elzen P., Tamaki S., Dunsmuir P. and Bedbrook J. 1985. Differential expression of the eight genes of the petunia ribulose biphosphate carboxylase small subunit multigene family. *EMBO J.* 4:3055-3061.

De Blas A.L. and Cherwinski H.M. 1983. Detection of antigens on nitrocellulose paper immunoblots with monoclonal antibodies. *Anal. Biochem.* 133:214-219.

Dehesh K., Tepperman J., Christensen A.H. and Quail P.H. 1991. *PhyB* is evolutionarily conserved and constitutively expressed in rice seedling shoots. *Mol. Gen. Genet.* 225:305-313.

De Lima D.C. 1974. *Isoenzymes cleaving S-cysteine derivatives from cauliflower.* M.S. Thesis. Univ. of California, Davis, California.

de Pater S., Hensgens L.A.M. and Schilperoort R.A. 1990. Structure and expression of a light-inducible shoot-specific rice gene. *Plant Mol. Biol.* 15:399-406.

Devereux J., Haeberli P. and Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387-395.

Dietz U.H., Gareis M. and Bertling W.M. 1990. Use of primer pools for specific gene amplification. *Nucleic Acids Res.* 18:5914.

Domec C., Garbay B., Fournier M. and Bonnet J. 1990. cDNA library construction from small amounts of unfractionated RNA: Association of cDNA synthesis with polymerase chain reaction amplification. *Anal. Biochem.* 188:422-426.

Dommisse E.M. 1993. *Development of an* Agrobacterium *transformation system for onion* (Allium cepa *L.*). Ph.D. thesis. Univ. of Canterbury, Christchurch, New Zealand.

Dommisse E.M., Leung D.W.M., Shaw M.L. and Conner A.J. 1990. Onion is a monocotyledonous host for *Agrobacterium*. *Plant Sci.* 69:249-257.

Draper J., Davey M.R., Freeman J.P., Cocking E.C. and Cox B.J. 1982. Ti Plasmid homologous sequences present in tissues from *Agrobacterium* plasmid-transformed *Petunia* protoplasts. *Plant & Cell Physiol.* 23:451-458.

foreign proteins in *Escherichia coli*. *Methods Enzymol.* 153:492-506.

Dugaiczyk A., Boyer H.W. and Goodman H.M. 1975. Ligation of *Eco*RI endonuclease-generated DNA fragments into linear and circular structures. *J. Mol. Biol.* 96:171-184.

Dunsmuir P. 1985. The petunia chlorophyll *a/b* binding protein genes: A comparison of *Cab* genes from different gene families. *Nucleic Acids Res.* 13:2503-2518.

Dunsmuir P., Smith S.M. and Bedbrook J. 1983. The major chlorophyll *a/b* binding protein of petunia is composed of several polypeptides encoded by a number of distinct nuclear genes. *J. Mol. Appl. Genet.* 2:285-300.

Dynan W.S. and Tjian R. 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* 316:774-778.

Edge A.S.B., Faltynek C.R., Hof L., Reichert L.E. Jr. and Weber P. 1981. Deglycosylation of glycoproteins by trifluoromethanesulfonic acid. *Anal. Biochem.* 118:131-137.

Edwards J.W. and Coruzzi G.M. 1990. Cell-specific gene expression in plants. *Ann. Rev. Genet.* 24:275-303.

Eguchi Y., Itoh T. and Tomizawa J.-I. 1991. Antisense RNA. *Ann. Rev. Biochem.* 60:631-652.

Ehrlich M. and Wang R.Y.-H. 1981. 5-Methylcytosine in eukaryotic DNA. *Science* 212:1350-1357.

Ericson M.L., Murén E., Gustavsson H.-O., Josefsson L.-G. and Rask L. 1991. Analysis of the promoter region of napin genes from *Brassica napus* demonstrates binding of nuclear protein *in vitro* to a conserved sequence motif. *Eur. J. Biochem.* 197:741-746.

Ettala T. and Virtanen A.I. 1962. Labeling of sulfur-containing amino acids and γ-glutamyl peptides after injection of labelled sulfate into onion. *Acta Chem. Scand.* 16:2061-2063.

Fang R.-X., Nagy F., Sivasubramaniam S. and Chua N.-H. 1989. Multiple *cis* regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. *The Plant Cell* 1:141-150.

Faye L. and Chrispeels M.J. 1988. Common antigenic determinants in the glycoproteins of plants, molluscs and insects. *Glycoconjugate J.* 5:245-256.

Feinberg A.P. and Vogelstein B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6-13.

Fenwick G.R. and Hanley A.B. 1985. The genus Allium - Part I. *CRC Crit. Rev. Food Sci. Nutr.* 22:199-271.

Fenwick G.R. and Hanley A.B. 1986. The genus Allium - Part II. *CRC Crit. Rev. Food Sci. Nutr.* 22:273-377.

Fenwick R.G. and Hanley A.B. 1990. Chemical composition. In: Rabinowitch H.D. and Brewster J.D., editors. *Onions and allied crops Vol. III. Biochemistry, food science and minor crops*, pages 17-32. CRC Press, Boca Raton, Florida.

Fischer E.H., Kent A.B., Snyder E.R. and Krebs E.G. 1958. The reaction of sodium borohydride with muscle phosphorylase. *J. Am. Chem. Soc.* 80:2906-2907.

Ford G.C., Eichele G. and Jansonius J.N. 1980. Three-dimensional structure of a pyridoxal phosphate-dependent enzyme, mitochondrial aspartate aminotransferase. *Proc. Natl. Acad. Sci. USA* 77:2559-2563.

Fowden L. 1964. The chemistry and metabolism of recently isolated amino acids. *Ann. Rev. Biochem.* 33:173-204.

Freeman G.G. 1975. Distribution of flavour components in onion (*Allium cepa* L.), leek (*Allium porrum*) and garlic (*Allium sativum*). *J. Sci. Food Agric.* 26:471-481.

Freeman G.G. 1979. Factors affecting flavour during growth, storage and processing of vegetables. In: Land D,G. and Nursten H.E., editors. *Progress in flavour research,* pages 225-243. Applied Science Publishers, Barking, Essex.

Freeman G.G. and Whenham R.J. 1975. A survey of volatile components of some *Allium* species in terms of *S*-alk(en)yl-L-cysteine sulphoxides present as flavour precursors. *J. Sci. Food Agric.* 26:1869-1886.

Freeman G.G. and Whenham R.J. 1976a. Thiopropanal *S*-oxide, alk(en)yl thiosulphinates and thiosulphonates: Simulation of flavour components of *Allium* species. *Phytochem.* 15:187-190.

Freeman G.G. and Whenham R.J. 1976b. Synthetic *S*-alk(en)yl-L-cysteine sulphoxides - alliinase fission products: Simulation of flavour components of *Allium* species. *Phytochem.* 15:521-523.

Freeman G.G., Whenham R.J., Mackenzie I.A. and Davey M.R. 1974. Flavour components in tissue cultures of onion (*Allium cepa* L.). *Plant Sci. Lett.* 3:121-125.

Fujita M., Endo M. and Sano M. 1990. Purification and characterization of alliin lyase from welsh onion, *Allium fistulosum* L. *Agric. Biol. Chem.* 54:1077-1079.

Fujiwara M., Yoshimura M., Tsuno S. and Murakami F. 1958. "Allithiamine", a newly found derivative of vitamin B_1. IV. On the alliin homologues in the vegetables. *J. Biochem.* 45:141-149.

Fütterer J. and Hohn T. 1992. Role of an upstream open reading frame in the translation of polycistronic mRNAs in plant cells. *Nucleic Acids Res.* 20:3851-3857.

Garnier J., Osguthorpe D.J. and Robson B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120:97-120.

Gasser, C.S., Winter J.A., Hironaka C.M. and Shah D.M. 1988. Structure, expression, and evolution of the 5-enolpyruvylshikimate-3-phosphate synthase genes of petunia and tomato. *J. Biol. Chem.* 263:4280-4289.

Protein Eng. 3:433-442.

Gilmartin P.M., Sarokin L., Memelink J. and Chua N.-H. 1990. Molecular light switches for plant genes. *The Plant Cell* 2:369-378.

Giovanelli J., Mudd S.H. and Datko A.H. 1980. Sulfur amino acids in plants. In: Miflin B.J. editor. *The biochemistry of plants - a comprehensive treatise Vol 5. Amino acids and derivatives*, pages 453-505. Academic Press, New York.

Girgis S.I., Alevizaki M., Denny P., Ferrier G.J.M. and Legon S. 1988. Generation of DNA probes for peptides with highly degenerate codons using mixed primer PCR. *Nucleic Acids Res.* 16:10371.

Giuliano G., Pichersky E., Malik V.S., Timko M.P., Scolnik P.A. and Cashmore A.R. 1988. An evolutionally conserved protein binding sequence upstream of a plant light regulated gene. *Proc. Natl. Acad. Sci. USA.* 85:7089-7093.

Goldberg R.B. 1986. Regulation of plant gene expression. *Phil. Trans. R. Soc. Lond. B* 314:343-353.

Goldberg R.B. 1988. Plants: Novel developmental processes. *Science* 240:1460-1467.

Goldberg R.B., Barker S.J. and Perez-Grau L. 1989. Regulation of gene expression during plant embryogenesis. *Cell* 56:149-160.

Gonzalez G.A., Yamamoto K.K., Fischer W.H., Karr D., Menzel P., Biggs III W., Vale W.W. and Montminy M.R. 1989. A cluster of phosphorylation sites on the cyclic AMP-regulated nuclear factor CREB predicted by its sequence. *Nature* 337:749-752.

Goodall G.J. and Filipowicz W. 1989. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* 58:473-483.

Goodall G.J. and Filipowicz W. 1990. The minimum functional length of pre-mRNA introns in monocots and dicots. *Plant Mol. Biol.* 14:727-733.

Goodall G.J. and Filipowicz W. 1991. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. *EMBO J.* 10:2635-2644.

Goodwin T.W. and Mercer E.I. 1983. *Introduction to plant biochemistry*, pages 227-272. Pergamon Press, Oxford.

Goore M.Y. and Thompson J.F. 1967. γ-Glutamyl transpeptidase from kidney bean fruit. I. Purification and mechanism of action. *Biochim. Biophys. Acta* 132:15-26.

Gould S.J., Subramani S. and Scheffler I.E. 1989. Use of the DNA polymerase chain reaction for homology probing: Isolation of partial cDNA or genomic clones encoding the iron-sulfur protein of succinate dehydrogenase from several species. *Proc. Natl. Acad. Sci. USA* 86:1934-1938.

Graham M.W., Doherty J.P. and Woodcock D.M. 1990. Efficient construction of plant genomic libraries requires the use of *mcr-* host strains and packaging mixes. *Plant Mol. Biol. Rep.* 8:18-27.

Granroth B. 1970. Biosynthesis and decomposition of cysteine derivatives in onion and other *Allium* species. *Ann. Acad. Sci. Fenn. Ser.* A2.154:1-71.

Green P.J., Kay S.A. and Chua N.-H. 1987. Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the *rbcS-3A* gene. *EMBO J.* 6:2543-2549.

Green P.J., Yong M.-H., Cuozzo M., Kano-Murakami Y., Silverstein P., and Chua N.-H. 1988. Binding site requirements for pea nuclear protein factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the *rbcS-3A* gene. *EMBO J.* 7:4035-4044.

Hall D.I. and Smith I.K. 1983. Partial purification and characterization of cystine lyase from cabbage (*Brassica oleracea* var *capitata*). *Plant Physiol.* 72:654-658.

Hamamoto A. and Mazelis M. 1986. The C-S lyases of higher plants. Isolation and properties of homogeneous cystine lyase from broccoli (*Brassica oleracea* var *botrytis*) buds. *Plant Physiol.* 80:702-706.

Hanelt P. 1990. Taxonomy, evolution and history. In: Rabinowitch H.D. and Brewster J.D., editors. *Onions and allied crops Vol. I. Botany, physiology and genetics*, pages 1-26. CRC Press, Boca Raton, Florida.

Hanley B.A. and Schuler M.A. 1988. Plant intron sequences: Evidence for distinct groups of introns. *Nucleic Acids Res.* 16:7159-7176.

Haqqi T.M., Sarkar G., David C.S. and Sommer S.S. 1988. Specific amplification with PCR of a refractory segment of genomic DNA. *Nucleic Acids Res.* 16:11844.

Harada J.J., Barker S.J. and Goldberg R.B. 1989. Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *The Plant Cell* 1:415-425.

Hatfield D.L., Smith D.W.E., Lee B.J., Worland P.J. and Oroszlan S. 1990. Structure and function of suppressor tRNAs in higher eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* 25:71-96.

Hattori T. and Nakamura K. 1988. Genes coding for the major tuberous root protein of sweet potato: Identification of putative regulatory sequence in the 5' upstream sequence. *Plant Mol. Biol.* 11:417-426.

Hauptmann R.M., Ozias-Akins P., Vasil V., Tabaeizadeh Z., Rogers S.G., Horsch R.B., Vasil I.K. and Fraley R.T. 1987. Transient expression of electroporated DNA in monocotyledonous and dicotyledonous species. *Plant Cell Rep.* 6:265-270.

Havey M.J. 1992. Restriction enzyme analysis of the nuclear 45s ribosomal DNA of six cultivated *Alliums* (*Alliaceae*). *Plant Syst. Evol.* 181:45-55.

Heery D.M., Gannon F. and Powell R. 1990. A simple method for subcloning DNA fragments from gel slices. *TIG.* 6:173.

Heidecker G. and Messing J. 1986. Structural analysis of plant genes. *Ann. Rev. Plant Physiol.* 37:439-466.

Heim U., Schubert R., Bäumlein H. and Wobus U. 1989. The leguminin gene family: Structure and evolutionary implications of *Vicia faba* B-type genes and pseudogenes. *Plant Mol. Biol.* 13:653-663.

Helfman D.M. and Hughes S.H. 1987. Use of antibodies to screen cDNA expression libraries prepared in plasmid vectors. *Methods Enzymol.* 152:451-469.

Herrmann A., Schulz W. and Hahlbrock K. 1988. Two alleles of the single-copy chalcone synthase in parsley differ by a transposon-like element. *Mol. Gen. Genet.* 212:93-98.

Hershey H.P., Barker R.F., Idler K.B., Lissemore J.L. and Quail P.H. 1985. Analysis of cloned cDNA and genomic sequences for phytochrome: Amino acid sequences for two gene products expressed in etiolated *Avena*. *Nucleic Acids Res.* 13:8543-8559.

Hightower R.C. and Meagher R.B. 1985. Divergence and differential expression of soybean actin genes. *EMBO J.* 4:1-8.

Holwerda B.C., Padgett H.S. and Rogers J.C. 1992. Proaleurin vacuolar targeting is mediated by short contiguous peptide interactions. *The Plant Cell* 4:307-318.

Hood L., Campbell J.H. and Elgin S.C.R. 1975. The organization, expression and evolution of antibody genes and other multigene families. *Ann. Rev. Genet.* 9:305-353.

Hooykaas P.J.J. 1989. Transformation of plant cells via *Agrobacterium*. *Plant Mol. Biol.* 13:327-336.

Horsch R.B., Fry J.E., Hoffmann N.L., Eichholtz D., Rogers S.G. and Fraley R.T. 1985. A simple and general method for transferring genes into plants. *Science* 227:1229-1231.

Hsu L.C., Okamoto M. and Snell E.E. 1989. L-Histidinol phosphate aminotransferase from *Salmonella typhimurium*. Kinetic behavior and sequence at the pyridoxal-P binding site. *Biochimie* 71:477-489.

Hung T., Mak K. and Fong K. 1990. A specificity enhancer for polymerase chain reaction. *Nucleic Acids Res.* 18:4953.

Hutchinson J. 1973. *The families of flowering plants*, pages 789-794. Oxford University Press, Oxford.

Huynh T.V., Young R.A. and Davis R.W. 1986. Constructing and screening cDNA libraries in λgt10 and λgt11. In: Glover D., editor. *DNA cloning techniques - a practical approach*, pages 49-78. IRL Press, Oxford.

Ingelbrecht I.L.W., Herman L.M.F., Dekeyser R.A., Van Montagu M.C. and Depicker A.G. 1989. Different 3' end regions strongly influence the level of gene expression in plant cells. *The Plant Cell* 1:671-680.

Iwami K. and Yasumoto K. 1980. Alliinase-like enzymes in fruiting bodies of *Lentinus edodes*: Their purification and substrate specificity. *Agric. Biol. Chem.* 44:3003-3004.

Jackson R.J. and Standart N. 1990. Do the poly(A)tail and the 3' untranslated region control mRNA translation? *Cell* 62:15-24.

Jacobsen J.V., Yamaguchi Y., Mann L.K. and Howard F.D. 1968. An alkyl-cysteine sulfoxide lyase in *Tulbaghia violacea* and its relation to other alliinase-like enzymes. *Phytochem.* 7:1099-1108.

Jansen H., Müller B. and Knobloch K. 1989a. Characterization of an alliin lyase preparation from garlic (*Allium sativum*). *Planta Med.* 55:434-439.

Jansen H., Müller B. and Knobloch K. 1989b. Alliin lyase from garlic, *Allium sativum*: Investigations on enzyme/substrate, enzyme/inhibitor interactions and on a new coenzyme. *Planta Med.* 55:440-445.

Joshi C.P. 1987a. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.* 15:6643-6653.

Joshi C.P. 1987b. Putative polyadenylation signals in nuclear genes of higher plants: A compilation and analysis. *Nucleic Acids Res.* 15:9627-9640.

Kaiser K. and Murray N.E. 1986. The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries. In: Glover D.M., editor. *DNA cloning techniques - a practical approach,* pages 1-47. IRL Press, Oxford.

Kajimura Y., Krull J., Miyakoshi S., Itakura K. and Toyoda H. 1990. Application of long synthetic oligonucleotides for gene analysis: Effect of probe length and stringency conditions on hybridisation specificity. *Genet. Anal. Techn. Appl.* 7:71-79.

Kamitani H., Esaki N., Tanaka H., Imahara H. and Soda K. 1990. Purification and characterization of *S*-alkylcysteine α,β-lyase from *Pseudomonas putida*. *J. Nutr. Sci. Vitaminol.* 36:339-347.

Kamitani H., Esaki N., Tanaka H. and Soda K. 1991. Degradation of L-djenkolate catalyzed by *S*-alkylcysteine α,β-lyase from *Pseudomonas putida*. *J. Biochem.* 109:645-649.

Kasai T. and Larson P.O. 1980. Chemistry and biochemistry of γ-glutamyl derivatives from plants including mushrooms (Basidiomycetes). In: Herz W., Grisebach H. and Kirby G.W., editors. *Progress in the chemistry of natural products*, pages 173-285, Springer Verlag, New York.

Kaulen H., Schell J. and Kreuzaler F. 1986. Light-induced expression of the chimeric chalcone synthase-NPTII gene in tobacco cells. *EMBO J.* 5:1-8.

Kay R., Chan A., Daly M. and McPherson J. 1987. Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. *Science* 236: 1299-1302.

Kay S.A., Keith B., Shinozaki K. and Chua N.-H. 1989. The sequence of the rice phytochrome gene. *Nucleic Acids Res.* 17:2865-2860.

Kazaryan R.A. and Goryachenkova V. 1978. Alliinase: Purification and characterization. *Biochimiya* 43:1905-1913.

Kean E.A. and Hare E.R. 1980. γ-Glutamyl transpeptidase of the ackee plant. *Phytochem.* 19:199-203.

Kirk J.T.O., Rees H. and Evans G. 1970. Base composition of nuclear DNA within the genus *Allium*. *Heredity* 25:507-512.

Klee H., Horsch R. and Rogers S. 1987. *Agrobacterium*-mediated plant transformation and its further applications to plant biology. *Ann. Rev. Plant Physiol.* 38:467-486.

Klein P., Kanehisa M. and DeLisi C. 1985. The detection and classification of membrane-spanning proteins. *Biochim. Biophys. Acta* 815:468-476.

Klein T.M., Wolf E.D., Wu R. and Sanford J.C. 1987. High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* 327:70-73.

Klemsdal S.S., Hughes W., Lönneborg A., Aalen R.B. and Olsen O.-A. 1991. Primary structure of a novel barley gene differentially expressed in immature aleurone layers. *Mol. Gen. Genet.* 228:9-16.

Kloeckener-Gruissem B., Vogel J.M. and Freeling M. 1992. The TATA box promoter region of maize *Adh1* affects its organ-specific expression. *EMBO J.* 11:157-166.

Knoth K., Roberds S., Poteet C. and Tamkun M. 1988. Highly degenerate, inosine-containing primers specifically amplify rare cDNA using the polymerase chain reaction. *Nucleic Acids Res.* 16:10932.

Koes R.E., Spelt C.E., Reif H.J., van den Elzen P.J.M., Veltkamp E. and Mol J.N.M. 1986. Floral tissue of *Petunia hybrida* (V30) expresses only one member of the chalcone synthase multigene family. *Nucleic Acids Res.* 14:5229-5239.

Kornfeld R. and Kornfeld S. 1985. Assembly of asparagine-linked oligosaccharides. *Ann. Rev. Biochem.* 54:631-664.

Kozak M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* 9:5233-5252.

Kozak M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857-872.

Kozak M. 1986. Bifunctional messenger RNAs in eukaryotes. *Cell* 47:481-483.

Kozak M. 1987. Effect of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Mol. Cell. Biol.* 7:3438-3445.

Kuhlemeier C. 1992. Transcriptional and post-transcriptional regulation of gene expression in plants. *Plant Mol. Biol.* 19:1-14.

Kuhlemeier C., Cuozzo M., Green P.J., Goyvaerts E., Ward K. and Chua N.-H. 1988. Localisation and conditional redundancy of regulatory elements in *rbcS-3A*, a pea gene encoding the small subunit of ribulose-phosphate carboxylase. *Proc. Natl. Acad. Sci. USA* 85:4662-4666.

Kuhlemeier C., Green P.J. and Chua N.-H. 1987. Regulation of gene expression in higher plants. *Ann. Rev. Plant Physiol.* 38:221-257.

Kyte J. and Doolittle R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105-132.

Laemmli U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 277:680-685.

Lam E. and Chua N.-H. 1989. ASF-2: A factor that binds to the cauliflower mosaic virus 35S promoter and a conserved GATA motif in *Cab* promoters. *The Plant Cell* 1:1147-1156.

Lancaster J.E. and Boland M.J. 1990. Flavor biochemistry. In: Rabinowitch H.D. and Brewster J.D., editors. *Onions and allied crops Vol. III. Biochemistry, food science, and minor crops,* pages 33-72. CRC Press, Boca Raton, Florida.

Lancaster J.E. and Collin H.A. 1981. Presence of alliinase in isolated vacuoles and of alkyl cysteine sulphoxides in the cytoplasm of bulbs of onion (*Allium cepa*). *Plant Sci. Lett.* 22:169-176.

Lancaster J.E., Dommisse E.M. and Shaw M.L. 1988a. Production of flavour precursors [*S*-alk(en)yl-L-cysteine sulphoxides] in photomixotrophic callus of garlic. *Phytochem.* 27:2123-2124.

Lancaster J.E., McCallion B.J. and Shaw M.L. 1984. The levels of *S*-alk(en)yl-L-cysteine sulphoxides during the growth of the onion (*Allium cepa* L.). *J. Sci. Food Agric.* 35:415-420.

Lancaster J.E., McCallion B.J. and Shaw M.L. 1986. The dynamics of the flavour precursors, the *S*-alk(en)yl-L-cysteine sulphoxides, during leaf blade and scale development in the onion (*Allium cepa*). *Physiol. Plant.* 66:293-297.

Lancaster J.E., Reynolds P.H.S., Shaw M.L., Dommisse E.M. and Munro J. 1988b. Intra-cellular localization of the biosynthetic pathway to flavour precursors in onion. *Phytochem.* 28:461-464.

Lancaster J.E. and Shaw M.L. 1989. γ-Glutamyl peptides in the biosynthesis of *S*-alk(en)yl-L-cysteine sulphoxides (flavour precursors) in *Allium*. *Phytochem.* 28:455-460.

Lassner M.W., Peterson P. and Yoder J.I. 1989. Simultaneous amplification of multiple DNA fragments by polymerase chain reaction in the analysis of transgenic plants and their progeny. *Plant Mol. Biol. Rep.* 7:116-128.

Last D.I., Brettell R.I.S., Chamberlain D.A., Chaudhury A.M., Larkin P.J., Marsh E.L., Peacock W.J. and Dennis E.S. 1991. pEmu: An improved promoter for gene expression in cereal cells. *Theoret. Appl. Genet.* 81:581-588.

Lawson L.D. and Hughes B.G. 1991. Characterization of the formation of allicin and other thiosulfinates from garlic. *Planta Med.* 58:345-350.

Lee C.C. and Caskey T. 1989. Generation of cDNA probes by reverse translation of amino acid sequence. *Genet. Eng.* 11:159-170.

Leeuw W.J.F., Slagboom P.E. and Vijg J. 1989. Quantitative comparison of mRNA levels in mammalian tissues: 28S ribosomal RNA level as an accurate internal control. *Nucleic Acids Res.* 17:10137-10138.

Lewin B. 1990. Commitment and activation at Pol II promoters: A tail of protein-protein interactions. *Cell* 61:1161-1164.

Leyva A., Liang X., Pintor-Toro J.A., Dixon R.A. and Lamb C.J. 1992. *cis*-Element combinations determine phenylalanine ammonia-lyase gene tissue-specific expression patterns. *The Plant Cell* 4:263-271.

Li W.-H., Luo C.-C. and Wu C.-I. 1985. Evolution of DNA sequences. In: MacIntyre R.J., editor. *Molecular evolutionary genetics*, pages 1-94. Plenum Press, New York.

Loake G.T., Faktor O., Lamb C.J. and Dixon R.A. 1992. Combination of H-box [CCTACC$(N)_7$CT] and G-box [CACGTG] *cis* elements is necessary for feed-forward stimulation of a chalcone synthase promoter by the phenylpropanoid-pathway intermediate ρ-coumaric acid. *Proc. Natl. Acad. Sci. USA* 89:9230-9234.

Logemann J., Schell J. and Willmitzer L. 1987. Improved method for the isolation of RNA from plant tissues. *Anal. Biochem.* 163:16-20.

Lois R., Dietrich A., Hahlbrock K. and Schulz W. 1989. A phenylalanine ammonia-lyase gene from parsley: Structure, regulation and identification of elicitor and light responsive *cis*-acting elements. *EMBO J.* 8:1641-1648.

Luehrsen K.R. and Walbot V. 1991. Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. *Mol. Gen. Genet.* 225:81-93.

Lütcke H.A., Chow K.C., Mickel F.S., Moss K.A., Kern H.F. and Scheele G.A. 1987. Selection of AUG initiation codons differs in plants and animals. *EMBO J.* 6:43-48.

MacDonald M.H., Mogen B.D. and Hunt A. 1991. Characterization of the polyadenylation signal from the T-DNA-encoded octopine synthase gene. *Nucleic Acids Res.* 19:5575-5581.

Malpathak N.P. and David S.B. 1986. Flavor formation in tissue cultures of garlic (*Allium sativum* L.). *Plant Cell Rep.* 5:446-447.

Martinez P., Martin W. and Cerff R. 1989. Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. *J. Mol. Biol.* 208:551-565.

Matikkala E.J. and Virtanen A.I. 1965a. γ-Glutamyl peptidase (glutaminase) in germinating seeds of chive (*Allium schoenoprasum*). *Acta Chem. Scand.* 19:1258-1261.

Matikkala E.J. and Virtanen A.I. 1965b. γ-Glutamyl peptidase in sprouting onion bulbs. *Acta Chem. Scand.* 19:1261-1262.

Matsuoka K. and Nakamura K. 1991. Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. *Proc. Natl. Acad. Sci. USA* 88:834-838.

Mazelis M. 1963. Demonstration and characterization of cysteine sulfoxide lyase in the Cruciferae. *Phytochem.* 2:15-22.

Mazelis M. and Creveling R.K. 1975. Purification and properties of *S*-alkyl-L-cysteine lyase from seedlings of *Acacia farnesiana* Willd. *Biochem. J.* 147:485-491.

Mazelis M. and Crews L. 1968. Purification of the alliin lyase of garlic, *Allium sativum* L. *Biochem. J.* 108:725-730.

Mazelis M. and Fowden L. 1973. Relationship of endogenous substrate to specificity of *S*-alkyl cysteine lyases of different species. *Phytochem.* 12:1287-1289.

McCallion B.J. and Lancaster J.E. 1984. Changes in the content and distribution, in different organs, of the flavour precursors, the *S*-alk(en)yl-L-cysteine sulphoxides, during seedling development of onions (*Allium cepa*) grown under light and dark regimes. *Physiol. Plant.* 62:370-374.

McElroy D., Rothenberg M., Reece K.S. and Wu R. 1990a. Characterization of the rice (*Oryza sativa*) actin gene family. *Plant Mol. Biol.* 15:257-268.

McElroy D., Rothenberg M. and Wu R. 1990b. Structural characterization of a rice actin gene. *Plant Mol. Biol.* 14:163-171.

McElroy D., Zhang W., Cao J. and Wu R. 1990c. Isolation of an efficient actin promoter for use in rice transformation. *The Plant Cell* 2:163-171.

McLauchlan J., Gaffney D., Whitton J.L. and Clements J.B. 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res.* 13:1347-1358.

McPherson K.G., Porter C. and Allison J.C.H. 1992a. Crop profiles - a database of significant crops grown in New Zealand. *CropSeed Report* 2.

McPherson M.J., Jones K.M. and Gurr S.J. 1992b. PCR with highly degenerate primers. In: McPherson M.J., Quirke. and Taylor G.R., editors. *PCR - a practical approach,* pages 171-186. Oxford University Press, Oxford.

Meinkoth J. and Wahl G. 1984. Hybridization of nucleic acids immobilized on solid supports. *Anal. Biochem.* 138:267-284.

Messing J. New M13 vectors for cloning. 1983. *Methods Enzymol.* 101:20-78.

Messing J., Geraghty D., Heidecker G., Hu N.-T., Kridl J. and Rubenstein I. 1986. Plant gene structure. In: Gustafson.J.P., editor. *Gene manipulation in plant improvement,* pages 211-227. Sixteenth Stadler Genetics, Plenum Press, New York.

Mignery G.A., Pikaard C.S. and Park W.D. 1988. Molecular characterization of the patatin multigene family of potato. *Gene* 62:27-44.

Mogen B.D., MacDonald M.H., Graybosch R. and Hunt A.G. 1990. Upstream sequences other than AAUAAA are required for efficient messenger RNA 3'-end formation in plants. *The Plant Cell* 2:1261-1272.

Mogen B.D., MacDonald M.H., Leggewie G. and Hunt A.G. 1992. Several distinct types of sequence elements are required for efficient mRNA 3' end formation in a pea *rbcS* gene. *Mol. Cell. Biol.* 12:5406-5414.

Mohanty B.D., Singh N.S. and Patwardhan M.V. 1989. Cytological changes associated with allicin production in garlic (*Allium sativum* L.) cultures. *Cell Chromosome Res.* 12:13-16.

Montell C., Fisher E.F., Caruthers M.H. and Berk A.J. 1983. Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' consensus sequence AAUAAA. *Nature* 305:600-605.

Moreno J. and Chrispeels M.J. 1989. A lectin gene encodes the α-amylase inhibitor of the common bean. *Proc. Natl. Acad. Sci. USA* 86:7885-7889.

Mullis K.B. and Faloona F.A. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155:335-350.

Murakami F. 1960. The nutritional value of allium plants XXXVI. Decomposition of alliin homologs by microorganisms and formation of substance with thiamide-masking activity. *Bitamin* 20:126-131;*Chem. Abstr.* 61:15094d (1964).

Murray E.E., Lotzer J. and Eberle M. 1989. Codon usage in plant genes. *Nucleic Acids Res.* 17:477-497.

Nash J., Luehrsen K.R. and Walbot V. 1990. *Bronze-2* gene of maize: Reconstruction of a wild-type allele and analysis of transcription and splicing. *The Plant Cell* 2:1039-1049.

Neuhaus J.-M., Flores S., Keefe D., Ahl-Goy P. and Meins F. Jr. 1992. The function of vacuolar β-1,3-glucanase investigated by antisense transformation. Susceptibility of transgenic *Nicotiana sylvestris* plants to *Cercospora nicotianae* infection. *Plant Mol. Biol.* 19:803-813.

Nevins J.R. 1983. The pathway of eukaryotic mRNA formation. *Ann. Rev. Biochem.* 52:441-466.

Nielsen N.C., Dickinson C.D., Cho T.-J., Thanh V.H., Scallon B.J., Fischer R.L., Sims T.L., Drews G.N. and Goldberg R.B. 1989. Characterization of the glycinin gene family in soybean. *The Plant Cell* 1:313-328.

Nock L.P. and Mazelis M. 1986. The C-S lyases of higher plants: Preparation and properties of homogeneous alliin lyase from garlic (*Allium sativum*). *Arch. Biochem. Biophys.* 249:27-33.

Nock L.P. and Mazelis M. 1987. The C-S lyases of higher plants. Direct comparison of the physical properties of homogeneous alliin lyase of garlic (*Allium sativum*) and onion (*Allium cepa*). *Plant Physiol.* 85:1079-1083.

Nock L.P. and Mazelis M. 1989. Lack of homology between the alliin lyases of garlic and onion. *Phytochem.* 28:729-731.

Nomura J., Nishizuka Y. and Hayaishi O. 1963. *S*-Alkylcysteinase: Enzymatic cleavage of *S*-methyl cysteine and its sulfoxide. *J. Biol. Chem.* 238:1441-1446.

Odell J.T., Nagy F. and Chua N.-H. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* 313:810-812.

Ohara O., Dorit R.L. and Gilbert W. 1989. One-sided polymerase chain reaction: The amplification of cDNA. *Proc. Natl. Acad. Sci. USA* 86:5673-5677.

Okamuro J.K. and Goldberg R.B. 1989. Regulation of plant gene expression: General principles. In: Marcus A., editor. *The biochemistry of plants - a comprehensive treatise, Vol. 15*, pages 1-82. Academic Press, New York.

Ow D.A., Jacobs J.D. and Howell S.H. 1987. Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporter of promoter activity. *Proc. Natl. Acad. Sci. USA* 84:4870-4874.

Paaren H.E., Slightom J.L., Hall T.C., Inglis A.S. and Blagrove R.J. 1987. Purification of a seed glycoprotein: *N*-terminal and deglycosylated analysis of phaseolin. *Phytochem.* 26:335-343.

Pain V.M. 1986. Initiation of protein synthesis in mammalian cells. *Biochem J.* 235:625-637.

Paterson D.R. 1979. Sulfur fertilization effects on onion yield and pungency. *Tex. Agric. Exp. Stn. Prog. Report* 3551.

Patil R.V. and Dekker E.E. 1990. PCR amplification of an *Escherichia coli* gene using mixed primers containing deoxyinosine at ambiguous positions in degenerate amino acid codons. *Nucleic Acids Res.* 18:3080.

Pearson L. and Meagher R.B. 1990. Diverse soybean actin transcripts contain a large intron in the 5' untranslated leader: Structural similarity to vertebrate muscle actin genes. *Plant Mol. Biol.* 14:513-526.

Pearson W.R. and Lipman D.J. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85:2444-2448.

Perlman D. and Halvorsen H.O. 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol.* 167:391-409.

Platenius H. 1944. Factors affecting onion pungency. *J. Agric. Res.* 62:371-379.

Poulsen C. and Chua N.-H. 1988. Dissection of 5' upstream sequences for selective expression of the *Nicotiana plumbaginifolia rbcS-8B* gene. *Mol. Gen. Genet.* 214:16-23.

Pugh B.F. and Tjian R. 1990. Mechanism of transcriptional activation by Sp1: Evidence for coactivators. *Cell* 61:1187-1197.

Pugsley A.P. 1989. *Protein targeting.* Academic Press, San Diego, California.

Raz V. and Fluhr R. 1992. Calcium requirement for ethylene-dependent responses. *The Plant Cell* 4:1123-1130.

Reddy A.R., Britsch L., Salamini F., Saedler H. and Wolfgang W. 1987. The A1 (anthocyanin-1) locus in *Zea mays* encodes dihydroquercitin reductase. *Plant Sci.* 52:7-13.

Reed K.C. and Mann D.A. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* 13:7207-7221.

Reynolds H.S., Smith L.A., Dickson J.M.J.J., Jones W.T., Rodber K.A., Carne A. and Liddane C.P. 1992. Molecular cloning of a cDNA encoding aspartate aminotransferase-P_2 from lupin nodules. *Plant Mol. Biol.* 19:465-472.

Rogers S.O. and Bendich A.J. 1988. Extraction of DNA from plant tissues. In: Gelvin S.B., Schilperoort R.A. and Verma D.P.S., editors. *Plant molecular biology manual*, chapter A6, pages 1-10. Kluwer Academic, Dordrecht.

Rolfe S.A. and Tobin E.M. 1991. Deletion analysis of a phytochrome-regulated monocot *rbcS* promoter in a transient assay system. *Proc. Natl. Acad. Sci. USA* 88:2683-2686.

Rothstein S.J., Dimaio J., Strand M. and Rice D. 1987. Stable and heritable inhibition of the expression of nopaline synthase in tobacco expressing antisense RNA. *Proc. Natl. Acad. Sci. USA* 84:8439-8443.

Rusche J.R. and Howard-Flanders P. 1985. Hexamine cobalt chloride promotes intermolecular ligation of blunt end DNA fragments by T4 DNA ligase. *Nucleic Acids Res.* 13:1997-2008.

Russel D.A. and Martin M.S. 1989. Differential expression and sequence analysis of the maize glyceraldehyde-3-phosphate dehydrogenase gene family. *The Plant Cell* 1:793-803.

Ryder T.B., Hedrick S.A., Bell J.N., Liang X., Clouse S.D. and Lamb C.J. 1987. Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. *Mol. Gen. Genet.* 210:219-233.

Saiki R.K. 1989. The design and optimization of the PCR. In: Erlich H.A., editor. *PCR technology - principles and applications for DNA amplification,* pages 7-16. Stockton Press, New York.

Sambrook J., Fritsch E.F. and Maniatis T. 1989. *Molecular cloning - A laboratory manual.* Second Edition. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Sanfaçon H. 1992. Regulation of mRNA formation in plants: Lessons from the cauliflower mosaic virus transcription signals. *Can. J. Bot.* 70:885-899.

Sanger F., Nicklen S. and Coulson A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.

Sarkar G., Kapelner S. and Sommer S.S. 1990. Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Res.* 18:7465.

Sato N. 1988. Nucleotide sequence and expression of the phytochrome gene in *Pisum sativum*: Differential regulation by light of multiple transcripts. *Plant Mol. Biol.* 11:697-710.

Schmid J., Doerner P.W., Clouse S.D., Dixon R.A. and Lamb C.J. 1990. Developmental and environmental regulation of a bean chalcone synthase promoter in transgenic tobacco. *The Plant Cell* 2:619-631.

Schulz W., Eiben H.-G. and Hahlbrock K. 1989. Expression in *Escherichia coli* of catalytically active phenylalanine ammonia-lyase from parsley. *FEBS Letts.* 258:335-338.

Schulze-Lefert P., Becker-André M., Schulz W., Hahlbrock K. and Dangl J.L. 1989a. Functional architecture of the light-responsive chalcone synthase promoter from parsley. *The Plant Cell* 1:707-714.

Schulze-Lefert P., Dangl J.L., Becker-André M., Hahlbrock K. and Schulz W. 1989b. Inducible *in vivo* DNA footprints define sequences necessary for UV light activation of the parsley chalcone synthase gene. *EMBO J.* 8:651-656.

Schwartz R.M. and Dayhoff M.O. 1978. Matrices for detecting distant relationships. In: Dayhoff M.O., editor. *Atlas of protein sequence and structure,* pages 353-358. National Biomedical Research Foundation, Washington.

Schwimmer S. 1969. Characterization of *S*-propenyl-L-cysteine sulfoxide as the principle endogenous substrate of L-cysteine sulfoxide lyase of onion. *Arch. Biochem. Biophys.* 130:312-320.

Schwimmer S. 1968. Enzymatic conversion of *trans*-(+)-*S*-1-propenyl-L-cysteine *S*-oxide to the bitter and odor-bearing components of onion. *Phytochem.* 7:401-404.

Schwimmer S. and Austin S.J. 1971. Gamma-glutamyl transpeptidase of sprouted onion. *J. Food Sci.* 36:807-811.

Schwimmer S. and Guadagni D.G. 1962. Relation between olfactory threshold concentration and pyruvic acid content of onion juice. *J. Food Sci.* 27:94-97.

Schwimmer S. and Guadagni D.G. 1968. Kinetics of the enzymatic development of pyruvic acid and odor in frozen onions treated with cysteine C-S lyase. *J. Food Sci.* 33:193-196.

Schwimmer S. and Kjaær A. 1960. Purification and specificity of the C-S-lyase of *Albizzia lophanta*. *Biochim. Biophys. Acta* 42:316-324.

Schwimmer S. and Mazelis M. 1963. Characterization of alliinase of *Allium cepa* (onion). *Arch. Biochem. Biophys.* 100:66-73.

Schwimmer S., Ryan C.A. and Wong F.F. 1964. Specificity of L-cysteine sulfoxide lyase and partially competitive inhibition by *S*-alkyl-L-cysteines. *J. Biol. Chem.* 239:777-782.

Schwimmer S. and Weston W.J. 1961. Enzymatic development of pyruvic acid in onions as a measure of pungency. *J. Ag. Food Chem.* 9:301-304.

Sebastiani F.L., Farrell L.B., Vasquez M. and Beachy R.N. 1991. Conserved amino acid sequences among plant proteins sorted to protein bodies and plant vacuoles. *Eur. J. Biochem.* 199:441-450.

Selby C., Galpin I.J. and Collin H.A. 1979. Comparison of the onion plant (*Allium cepa*) and onion tissue culture. *New Phytol.* 83:351-359.

Serfling E., Jasin M. and Schaffner W. 1985. Enhancers and eukaryotic gene transcription. *TIG.* 1:224-230.

Sharp P.A. 1987. Splicing of messenger RNA precursors. *Science* 235:766-771.

Sharrock R.A., Lissemore J.L. and Quail P.H. 1986. Nucleotide and amino acid sequence of a *Cucurbita* phytochrome cDNA clone: Identification of conserved features by comparison with *Avena* phytochrome. *Gene* 47:287-295.

Sharrock R.A. and Quail P.H. 1989. Novel phytochrome sequences in *Arabidopsis thaliana*: Structure, evolution, and differential expression of a plant regulatory photoreceptor family. *Genes Dev.* 3:1745-1757.

Shinshi H., Neuhaus J.-M., Ryals J. and Meins F. Jr. 1990. Structure of a tobacco endochitinase gene: Evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain. *Plant Mol. Biol.* 14:357-368.

Short J.M., Fernandez J.M., Sorge J.A. and Huse W.D. 1988. λZAP: A bacteriophage λ expression vector with *in vivo* excision properties. *Nucleic Acids Res.* 16:7583-7600.

Simon C.J. and Weeden N.F. 1992. Molecular analysis and cloning of *Malus* ribosomal DNA. *J. Amer. Soc. Hort. Sci.* 117:164-168.

Simpson J., Schell J., Van Montagu M. and Herrera-Estrella L. 1986a. Light-inducible and tissue-specific pea *lhcp* gene expression involves an upstream element combining enhancer- and silencer-like properties. *Nature* 323:551-554.

Simpson J., Van Montagu M. and Herrera-Estrella L. 1986b. Photosynthesis-associated gene families: Differences in response to tissue-specific and environmental factors. *Science* 233:34-38.

Slightom J.L. and Drong R.F. 1988. Procedures for constructing genomic clone banks. In: Gelvin S.B., Schilperoort, R.A. and Verma D.P.S., editors. *Plant molecular biology manual,* chapter A8, pages 1-52. Kluwer Academic, Dordrecht.

Smith C.J.S., Watson C.F., Ray J., Bird C.R., Morris P.C., Schuch W. and Grierson D. 1988. Antisense RNA inhibition of polygalacturonase gene expression in transgenic tomatoes. *Nature* 334:724-726.

Somers D.A., Rines H.W., Gu W., Kaeppler H.F. and Bushnell W.R. 1992. Fertile, transgenic oat plants. *Bio/Technol.* 10:1589-1594.

Sommer R. and Tautz D. 1989. Minimal homology requirements for PCR primers. *Nucleic Acids Res.* 17:6749.

Southern E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.

Stack S.M. and Comings D.E. 1978. The chromosomes and DNA of *Allium cepa. Chromosoma* 70:161-181.

Stoll A. and Seebeck E. 1947. Alliin, the pure mother substance of garlic oil. *Experientia* 3:114-115.

Stoll A. and Seebeck E. 1951. Chemical investigations on Alliin, the specific principle of garlic. *Adv. Enzymol.* 11:377-400.

Sugii M., Nagasawa S. and Suzuki T. 1963a. Biosynthesis of *S*-methyl-L-cysteine and *S*-methyl-L-cysteine sulfoxide from methionine in garlic. *Chem. Pharm. Bull.* 11:135-136;*Chem. Abstr.* 58:1447b (1963).

Sugii M., Suzuki T. and Nagasawa S. 1963b. Isolation of (-)-*S*-propenyl-L-cysteine from garlic. *Chem. Pharm. Bull.* 11:548-547;*Chem. Abstr.* 59:6509b (1963).

Suzuki T., Sugii M. and Kakimoto T. 1962. Metabolic incorporation of L-valine-C^{14} into *S*-(2-carboxypropyl)glutathione and *S*-(2-carboxypropyl)cysteine in garlic. *Chem. Pharm. Bull.* 10:328-331;*Chem. Abstr.* 59:11561b (1962).

Sweet W.J. and Mazelis M. 1987. Homogeneous alkylcysteine lyase of *Acacia farnesiana*: Fresh seedlings vs. acetone powders. *Phytochem.* 26:945-948.

Tague B.W. and Chrispeels M.J. 1987. The plant vacuolar protein, phytohemagglutinin, is transported to the vacuole of transgenic yeast. *J. Cell Biol.* 105:1971-1979.

Tague B.W., Dickinson C.D. and Chrispeels M.J. 1990. A short domain of the plant vacuolar protein phytohemagglutinin targets invertase to the yeast vacuole. *The Plant Cell* 2:533-546.

Takatsuji H., Mori M., Benfey P.N., Ren L. and Chua N.-H. 1991. Characterization of a zinc finger DNA-binding protein expressed specifically in petunia petals and seedlings. *EMBO J.* 11:241-249.

Takeda Y., Ohlendorf D.H. Anderson W.F. and Matthews B.W. 1983. DNA-binding proteins. *Science* 221:1020-1026.

Tanase S., Kojima H. and Morino Y. 1979. Pyridoxal 5'-phosphate binding site of pig heart alanine aminotransferase. *Biochem.* 18:3002-3007.

Tate W.P. and Caskey C.T. 1990. Termination of protein synthesis. In: Spedding G., editor. *Ribosomes and protein synthesis - a practical approach,* pages 81-100. IRL Press, Oxford.

Tesniere C. and Vayda M.E. 1991. Method for the isolation of high-quality RNA from grape berry tissues without contaminating tannins or carbohydrates. *Plant Mol. Biol. Rep.* 9:242-251.

Thach R.E. 1992. Cap recap: The involvement of eIF-4F in regulating gene expression. *Cell* 68:177-180.

Thomas P.S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* 77:5201-5205.

Thompson W.F. and White M.J. 1991. Physiological and molecular studies of light-regulated nuclear genes in higher plants. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 42:423-466.

Tigier H.A., Quesada M.A., Heredia A. and Valpuesta V. 1991. Partial deglycosylation of an anionic isoperoxidase from peach seeds - effect on enzyme activity, stability and antigenicity. *Physiol. Plant.* 83:144-148.

Tobin E.M. and Silverthorne J. 1985. Light regulation of gene expression in higher plants. *Ann. Rev. Plant Physiol.* 36:569-593.

Tobkin H.E. and Mazelis M. 1979. Alliin lyase: Preparation and characterization of the homogeneous enzyme from onion bulbs. *Arch. Biochem. Biophys.* 193:150-157.

Touchet N., Chardin P. and Tavitian A. 1987. Four additional members of the *ras* gene superfamily isolated by an oligonucleotide strategy: Molecular cloning of YPT-related cDNAs from a rat brain library. *Proc. Natl. Acad. Sci. USA* 84:8210-8214.

Tregear J.W. and Roberts L.M. 1992. The lectin gene family of *Ricinus communis*: Cloning of a functional ricin gene and three lectin pseudogenes. *Plant Mol. Biol.* 18:515-525.

Tsuno S. 1958. Alliinase in *Allium* plants. *Bitamin* 14:659-664.

Turnbull A., Galpin I.J., Smith J.L. and Collin H.A. 1981. Comparison of the onion plant (*Allium cepa*) and onion tissue culture IV. Effect of shoot and root morphogenesis on flavour precursor synthesis in onion tissue culture. *New Phytol.* 87:257-268.

Udvardi M.K. and Kahn M.L. 1991. Isolation and analysis of a cDNA clone that encodes an alfalfa (*Medicago sativa*) aspartate aminotransferase. *Mol. Gen. Genet.* 231:97-105.

van Damme E.J.M., Smeets K., Torrekens S., van Leuven F., Goldstein I.J. and Peumans W.J. 1992. The closely related homomeric and heterodimeric mannose-binding lectins from garlic are encoded by one-domain and two-domain lectin genes, respectively. *Eur. J. Biochem.* 206:413-420.

van der Krol A., Lenting P.E., Veenstra J., van der Meer I.M., Koes R.E., Gerats A.G.M., Mol J.N.M. and Stuitje A.R. 1988. An anti-sense chalcone synthase gene in transgenic plants inhibits flower pigmentation. *Nature* 333:866-869.

van der Meer I.M., Brouwer M., Spelt C.E., Mol J.N.M. and Stuitje A.R. 1992a. The TACPyAT repeats in the chalcone synthase promoter of *Petunia hybrida* act as a dominant negative *cis*-acting module in the control of organ-specific expression. *The Plant J.* 2:525-535.

van der Meer I.M., Stam M.E., van Tunen A.J., Mol J.N.M. and Stuitje A.R. 1992b. Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. *The Plant Cell* 4:253-262.

Virtanen A.I. 1965. Studies on organic sulphur compounds and other labile substances in plants. *Phytochem.* 4:207-228.

Virtanen A.I. and Matikkala E.J. 1959. The isolation of *S*-methylcysteine-sulphoxide and *S*-*n*-propylcysteinesulphoxide from onion (*Allium cepa*) and the antibiotic activity of crushed onion. *Acta. Chem. Scand.* 13:1898-1900.

Virtanen A.I. and Spåre C.-G. 1961. Isolation of the precursor of the lachrymatory factor in onion (*Allium cepa*). *Suom. Kemistil.* B34:72-74.

Voelker T.A., Moreno J. and Chrispeels M.J. 1990. Expression analysis of a pseudogene in transgenic tobacco: A frameshift mutation prevents mRNA accumulation. *The Plant Cell* 2:255-261.

von Heijne G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683-4690.

von Schaewen A., Stitt M., Schmidt R., Sonnewald U. and Willmitzer L. 1990. Expression of a yeast derived invertase in the cell wall of tobacco and *Arabidopsis* plants leads to accumulation of carbohydrate and inhibition of photosynthesis and strongly influences growth and phenotype of transgenic tobacco plants. *EMBO J.* 9:3033-3044.

Wada K., Wada Y., Ishibashi F., Gojobori T. and Ikemura T. 1992. Codon usage tabulated from the Genbank genetic sequence data. *Nucleic Acids Res.* 20: 2111-2118.

Wahba A.J., Woodley C.L. and Dholakia J.N. 1990. Initiation of protein synthesis. In: Spedding G., editor. *Ribosomes and protein synthesis - a practical approach,* pages 31-50. IRL Press, Oxford.

Wahle E. 1992. The end of the message: 3'-end processing leading to polyadenylated messenger RNA. *Bioessays* 14:113-118.

Wallace R.B. and Miyada C.G. 1987. Oligonucleotide probes for the screening of recombinant DNA libraries. *Methods Enzymol.* 152:432-442.

Wang Y., Zhang W., Cao J., McElroy D. and Wu R. 1992. Characterization of *cis*-acting elements regulating transcription from the promoter of a constitutively active rice actin gene. *Mol. Cell. Biol.* 12:3399-3406.

Wefald F.C., Devlin B.H. and Williams R.S. 1990. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. *Nature* 344:260-262.

Weisshaar B., Armstrong G.A., Block A., da Costa e Silva O. and Hahlbrock K. 1991. Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness. *EMBO J.* 10:1777-1786.

Welling G.W., Weijer W.J., van der Zee R. and Welling-Wester S. 1985. Prediction of sequential antigenic regions in proteins. *FEBS Lett.* 188:215-218.

Weretilnyk E.A. and Hanson A.D. 1990. Molecular cloning of a plant betaine-aldehyde dehydrogenase, an enzyme implicated in adaptation to salinity and drought. *Proc. Natl. Acad. Sci. USA* 87:2745-2749.

Whitaker J.R. 1976. Development of flavor, odor and pungency in onion and garlic. *Adv. Food Res.* 22:73-133.

White O., Soderlund C., Shanmugan P. and Fields C. 1992. Information contents and dinucleotide compositions of plant intron sequences vary with evolutionary origin. *Plant Mol. Biol.* 19:1057-1064.

Whitehead A.S., Goldberger G., Woods D.E., Markham A.F. and Colten H.R. 1983. Use of a cDNA clone for the fourth component of human complement (C4) for analysis of a genetic deficiency of C4 in guinea pig. *Proc. Natl. Acad. Sci. USA* 80:5387-5391.

Wickens M. and Stephenson P. 1984. Role of the conserved AAUAAA sequence: Four AAUAAA point mutations prevent messenger RNA 3' end formation. *Science* 226:1045-1051.

Wilkins T.A. and Raikhel N.V. 1989. Expression of rice lectin is governed by two temporally and spatially regulated mRNAs in developing embryos. *The Plant Cell* 1:541-549.

Wilkins T.A., Bednarek S.Y. and Raikhel N.V. 1990. Role of propeptide glycan in post-translational processing and transport of barley lectin to vacuoles in transgenic tobacco. *The Plant Cell* 2:301-313.

Wilkins W.F. 1961. *The isolation and identification of the lachrymogenic compounds of onion.* PhD Thesis. Cornell University, New York.

Williams M.E., Foster R. and Chua N.-H. 1992. Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. *The Plant Cell* 4:485-496.

Wozney J.M. 1990. Using purified protein to clone its gene. *Methods Enzymol.* 182:738-751.

Yamamoto K.R. 1985. Steroid receptor regulated transcription of specific genes and gene networks. *Ann. Rev. Genet.* 19:209-252.

Yip W.-K., Dong J.-G., Kenny J.W., Thompson G.A. and Yang S.F. 1990. Characterization and sequencing of the active site of 1-aminocyclopropane-1-carboxylate synthase. *Proc. Natl. Acad. Sci. USA* 87:7930-7934.

Zimmerman S.B. and Harrison B. 1985. Macromolecular crowding accelerates the cohesion of DNA fragments with complementary termini. *Nucleic Acids Res.* 13:2241-2249.

Appendix I

DNA and predicted amino acid sequence of the cDNA clone Alli4A

```
1 A AGG TCG TGT GAC GCA TCA TTT CGC TTT CTC TCT ATA AAA CTA   43

 44 TCT TTA CTC TTT ACA ATC TTA CTC TTA TTT TCT TCA AAA CCA   85

 86 ATC GCC ATG GCT ACT AAA TGT GGA GGC ATT CAT GAA AAG GAA  127
           MET Ala Thr Lys Cys Gly Gly Ile His Glu Lys Glu

128 GGA ACT GAG AAC AGT GTC GAG ATC GAA GAA CTC GCT CGA TTT  169
    Gly Thr Glu Asn Ser Val Glu Ile Glu Glu Leu Ala Arg Phe

170 GCT GTC GAC GAA CAC AAC AAA AAA GAG AAT GCA CTA TTG GAG  211
    Ala Val Asp Glu His Asn Lys Lys Glu Asn Ala Leu Leu Glu

212 TTT GGT CGT GTG CTG AAG GCG AAC GAA CAA GTG GTG GCT GGC  253
    Phe Gly Arg Val Leu Lys Ala Asn Glu Gln Val Val Ala Gly

254 ACC ATG CAT CAC TTG ACT GTA GAA GCA ATC GAG GCA GGG AAA  295
    Thr MET His His Leu Thr Val Glu Ala Ile Glu Ala Gly Lys

296 AAG AAG ATA TAC GAG GCC AAG GTT TGG GTT AAA CCA TGG CTT  337
    Lys Lys Ile Tyr Glu Ala Lys Val Trp Val Lys Pro Trp Leu

338 AAC TTT AAA GAA CTT CAG GAG TTC AGA CAT GCA GGA GAC TCT  379
    Asn Phe Lys Glu Leu Gln Glu Phe Arg His Ala Gly Asp Ser

380 ATT CCT AAT GCT ACT CCT TCA GAC CTT GGA GCT AAA CGG GGT  421
    Ile Pro Asn Ala Thr Pro Ser Asp Leu Gly Ala Lys Arg Gly

422 GAT CAC CCA CCT GGA TGG CGT GAT GTG CCT GCT CAT GAT CCA  463
    Asp His Pro Pro Gly Trp Arg Asp Val Pro Ala His Asp Pro

464 GTT GTG AAA GAT GCA GCA GAT CAT GCA GTG AAG ACA ATT CAA  505
    Val Val Lys Asp Ala Ala Asp His Ala Val Lys Thr Ile Gln

506 CAG AGG TCA AAT TCC CTT GCT GCT TAT GAA CTG CTT GAA GTT  547
    Gln Arg Ser Asn Ser Leu Ala Ala Tyr Glu Leu Leu Glu Val

548 CTT CAT GCC AAG GCT GAG GTG ATA GAT GAT CTT GCC AAG TTT  589
    Leu His Ala Lys Ala Glu Val Ile Asp Asp Leu Ala Lys Phe

590 GAC TTG CTT CTT AAA CTA AAG CGG GGA AAC AAG GAA GAG AAA  631
    Asp Leu Leu Leu Lys Leu Lys Arg Gly Asn Lys Glu Glu Lys

632 TAT AAA GTT GAA GTA CAC AAG AAT ATG GAA GGC AAC TTT CTT  673
    Tyr Lys Val Glu Val His Lys Asn MET Glu Gly Asn Phe Leu

674 CTG AAT CAG ATG CAA CAG GAG CAT TGA GCT GCT CCT GGC TGT  715
    Leu Asn Gln MET Gln Gln Glu His TER

716 GAA TAT TTG TAT CCA GTA AAA ACT TGA ATA TGT ATG TGT TTG  757

758 TGT GAA CTC TTA TTA CTG CCT TAC TAG TCT GAA ATA AAA TAG  799

800 TGC AGT AAA TAT GTT ATC TTA AAA AAA AAA AAA AAA AAA AAA  841
```

Appendix II
DNA and predicted amino acid sequence of the Alli9 extraneous cDNA

```
1 AC CCT AGA GTT TTC TTG ACA TGG AGG TCG GCG CCC CCC GCC GGT  44

  45 CGT ATT GTC ATG GAG CTA TTC CAA GAC GTT GTC CCA CAA ACA  86
                 MET Glu Leu Phe Gln Asp Val Val Pro Gln Thr

  87 GCC GAG AAC TTC CGT GCG CTC TGC ACA GGA GAG AAG GGC ATG 128
     Ala Glu Asn Phe Arg Ala Leu Cys Thr Gly Glu Lys Gly MET

 129 GGC GAT CGC AAG CCT CTC CAC TAC AAG GGA TCT AGC TTC CAT 170
     Gly Asp Arg Lys Pro Leu His Tyr Lys Gly Ser Ser Phe His

 171 CGG GTG ATC CCC GGT TTC ATG TGC CAA GGA GGC GAC TTT ACT 212
     Arg Val Ile Pro Gly Phe MET Cys Gln Gly Gly Asp Phe Thr

 213 GCT GGA AAT GGT ACC GGA GGT GAG TCG ATC TAC GGT GCT AAG 254
     Ala Gly Asn Gly Thr Gly Gly Glu Ser Ile Tyr Gly Ala Lys

 255 TTC AAG GAT GAG AAC TTC ATC AAA AAG CAC ACC GGA CCT GGT 296
     Phe Lys Asp Glu Asn Phe Ile Lys Lys His Thr Gly Pro Gly

 297 GTC TTG TCA ATG GCA AAC GCT GGA CCT GGA ACC AAT GGT TCT 338
     Val Leu Ser MET Ala Asn Ala Gly Pro Gly Thr Asn Gly Ser

 339 CAG TTT TTC ATC TGC ACT GAA AAG ACT GCC TGG CTT GAT GGC 380
     Gln Phe Phe Ile Cys Thr Glu Lys Thr Ala Trp Leu Asp Gly

 381 AAG CAT GTC GTT TTC GGA CAG GTT GTG GAA GGC ATG GAT GTT 422
     Lys His Val Val Phe Gly Gln Val Val Glu Gly MET Asp Val

 423 GTG CGA GCC ATT GAG AAG GTT GGA TCT CAG AGT GGA CAG ACC 464
     Val Arg Ala Ile Glu Lys Val Gly Ser Gln Ser Gly Gln Thr

 465 AAG AAA CCT GTT AAG ATT GCC GAT TGT GGT CAG CTT TCT TAG 506
     Lys Lys Pro Val Lys Ile Ala Asp Cys Gly Gln Leu Ser TER

 507 ATC AAT GCT TGT TTT TGT TAT TGC TTG AGT CTA TGT TTC TAA 548

 549 TAA TGA AAA CGT TAT GAG TTG TGT TAC TTT GTT TTC ATG TTT 590

 591 TAT GAT ATG AAT CTG AGT TAT TAA AAA AAA AAA AAA AAA A   630
```

Appendix III

Data generated from the alliinase amino acid sequence inferred from the cDNA clones

Hydrophilicity (Kyte-Doolittle) averaged over a window of 7 residues (HyPhil).
Surface Probability according to Emini (SurfPr).
Chain Flexibility according to Karplus-Schulz (FlexPr).
Secondary Structure according to Chou-Fasman (CF-Pred).
Secondary Structure according to Garnier-Osguthorpe-Robson (GORPred).
Antigenicity Index according to Jameson-Wolf (AI-Ind).

Pos	AA	Glycos	HyPhil	SurfPr	FlexPr	CF-Pred	GORPred	AI-Ind
1	M	.	0.925	1.348	1.000	.	H	0.900
2	E	.	1.380	1.435	1.000	.	H	0.900
3	S	.	1.800	2.245	1.000	.	H	0.900
4	Y	.	0.943	1.684	1.000	.	H	0.900
5	H	.	1.271	0.962	0.994	.	H	0.600
6	K	.	0.886	0.962	1.012	.	H	0.750
7	V	.	1.271	0.987	1.032	.	T	1.150
8	G	.	1.643	1.451	1.063	.	T	1.300
9	S	.	0.914	0.718	1.080	T	T	1.550
10	N	.	0.586	1.496	1.079	T	T	1.700
11	K	.	1.300	2.026	1.070	.	.	0.900
12	M	.	0.700	1.247	1.047	.	.	0.900
13	P	.	0.043	0.639	1.028	T	.	0.850
14	S	.	-1.100	0.224	1.001	T	.	-0.050
15	L	.	-2.200	0.187	0.971	B	B	-0.600
16	L	.	-2.571	0.085	0.942	B	B	-0.600
17	I	.	-3.157	0.034	0.916	B	B	-0.600
18	L	.	-3.914	0.029	0.903	B	B	-0.600
19	I	.	-4.014	0.024	0.894	B	B	-0.600
20	C	.	-3.743	0.035	0.891	B	B	-0.600
21	I	.	-2.986	0.056	0.895	B	B	-0.600
22	I	.	-2.329	0.107	0.910	B	B	-0.600
23	M	.	-2.086	0.173	0.930	B	B	-0.600
24	S	.	-2.329	0.184	0.956	B	B	-0.600
25	S	.	-1.186	0.421	0.971	B	.	-0.600
26	F	.	-0.043	0.684	0.979	B	.	-0.300
27	V	.	0.729	0.821	0.990	B	.	0.600
28	N	.	-0.029	0.430	0.996	t	.	-0.100
29	N	.	-0.400	0.501	1.000	t	.	-0.250
30	N	.	0.500	1.169	0.985	t	.	0.950
31	I	.	0.843	0.734	0.967	H	.	0.600
32	A	.	0.900	0.913	0.961	H	.	0.600
33	Q	.	-0.200	0.422	0.962	H	.	-0.300
34	A	.	-0.600	0.868	0.980	H	.	-0.600
35	K	.	0.171	0.903	0.988	H	.	0.300
36	V	.	0.543	0.699	0.978	H	H	0.600
37	T	.	-0.500	0.571	0.964	H	H	-0.600
38	W	.	0.314	0.571	0.955	H	H	0.300
39	S	.	-0.500	0.777	0.955	H	H	-0.600
40	L	.	-0.157	0.544	0.964	H	H	-0.300
41	K	.	0.243	0.896	0.979	H	H	0.300
42	A	.	0.614	1.157	0.985	H	H	0.750
43	A	.	0.243	1.418	0.995	H	H	0.450
44	E	.	1.286	1.228	1.003	H	H	0.900
45	E	.	0.471	1.228	1.002	H	H	0.600
46	A	.	0.129	0.902	0.993	H	H	0.300
47	E	.	0.129	0.526	0.973	H	H	0.300
48	A	.	0.129	0.489	0.954	H	H	0.300
49	V	.	-1.014	0.339	0.944	H	H	-0.600
50	A	.	-0.257	0.315	0.942	H	H	-0.300
51	N	.	-1.114	0.167	0.947	.	H	-0.600
52	I	.	-0.743	0.302	0.960	t	H	-0.400

53	N	G	-0.086	0.295	0.972	t	T	0.300
54	C	.	0.629	0.250	0.986	T	T	1.400
55	S	.	0.186	0.353	1.004	T	T	1.250
56	G	.	1.471	0.430	1.014	T	T	1.550
57	H	.	0.714	0.810	1.015	t	T	1.350
58	G	.	0.671	0.523	1.009	t	.	0.950
59	R	.	0.014	0.436	0.992	.	B	0.300
60	A	.	0.457	0.535	0.980	b	B	0.300
61	F	.	0.057	0.535	0.981	b	B	0.300
62	L	.	-0.643	0.192	0.984	b	B	-0.600
63	D	.	-1.829	0.156	0.993	b	B	-0.600
64	G	.	-1.457	0.242	0.995	b	B	-0.600
65	I	.	-0.557	0.490	0.990	b	B	-0.600
66	L	.	0.043	0.290	1.003	b	.	0.450
67	S	.	-0.343	0.393	1.028	T	T	0.650
68	D	.	-0.171	0.868	1.057	T	T	0.650
69	G	.	1.029	2.104	1.089	T	T	1.700
70	S	.	1.214	0.842	1.094	t	.	0.950
71	P	.	1.600	0.873	1.071	T	T	1.550
72	K	.	0.743	0.473	1.042	T	T	1.550
73	C	.	1.186	0.567	1.006	t	T	1.350
74	E	.	1.171	0.529	0.986	t	T	1.200
75	C	.	0.586	0.142	0.977	t	T	1.200
76	N	.	0.214	0.415	0.973	B	T	0.700
77	T	.	0.671	0.346	0.976	B	T	1.000
78	C	.	0.229	0.638	0.973	B	T	0.700
79	Y	.	0.329	0.401	0.984	B	T	0.700
80	T	.	0.329	0.464	0.996	B	T	0.700
81	G	.	-0.129	0.464	1.000	.	T	0.250
82	A	.	0.343	0.397	1.005	T	T	1.250
83	D	.	0.657	0.476	1.003	T	T	1.550
84	C	.	1.114	0.962	1.011	t	T	1.350
85	S	.	0.414	0.668	1.027	T	T	1.250
86	E	.	0.771	0.577	1.042	T	T	1.550
87	K	.	0.329	1.065	1.051	.	T	1.000
88	I	.	0.329	0.426	1.043	.	T	0.850
89	T	.	0.329	0.330	1.027	.	T	0.850
90	G	.	-0.429	0.167	1.011	t	T	0.150
91	C	.	-0.486	0.397	0.991	t	B	-0.400
92	S	.	-0.443	0.204	0.976	.	B	-0.600
93	A	.	-0.800	0.208	0.968	.	B	-0.600
94	D	.	-0.743	0.521	0.961	.	B	-0.600
95	V	.	-0.329	0.385	0.974	.	B	-0.300
96	A	.	0.057	0.636	0.996	t	B	0.500
97	S	.	0.371	0.377	1.022	t	.	0.650
98	G	.	-0.671	0.419	1.042	T	.	-0.050
99	D	.	-0.471	0.359	1.035	T	.	-0.050
100	G	.	-0.757	0.221	1.017	t	.	-0.250
101	L	.	-0.371	0.386	0.992	H	.	-0.300
102	F	.	0.071	0.401	0.978	H	.	0.300
103	L	.	-0.243	0.634	0.979	H	H	-0.300
104	E	.	-0.171	0.809	0.981	H	H	-0.300
105	E	.	0.871	1.618	0.984	H	H	0.750
106	Y	.	1.771	3.397	0.981	H	H	0.750
107	W	.	2.771	2.669	0.981	H	H	0.750
108	Q	.	2.829	3.083	0.993	H	H	0.750
109	Q	.	2.829	3.407	1.010	H	H	0.900
110	H	.	3.143	5.211	1.034	H	H	0.900
111	K	.	3.129	4.032	1.058	H	.	0.900
112	E	.	2.371	2.352	1.063	H	.	0.900
113	N	.	1.271	1.283	1.053	h	.	0.900
114	S	.	0.271	0.529	1.025	B	.	0.450
115	A	.	-0.886	0.227	0.985	B	.	-0.600
116	V	.	-1.271	0.189	0.961	B	.	-0.600
117	L	.	-1.714	0.140	0.955	B	.	-0.600
118	V	.	-1.700	0.145	0.962	B	.	-0.600
119	S	.	-0.986	0.266	0.978	.	.	-0.600
120	G	.	0.257	0.632	0.978	.	T	0.700

121	W	.	0.529	0.843	0.964	.	T	1.000
122	H	.	1.243	0.843	0.951	.	.	0.600
123	R	.	1.314	1.335	0.935	.	T	1.150
124	M	.	0.857	1.099	0.928	.	T	1.150
125	S	.	0.329	0.700	0.926	.	T	0.700
126	Y	.	0.371	0.574	0.929	B	T	0.700
127	F	.	-0.043	0.897	0.942	B	T	0.100
128	F	.	-0.371	0.497	0.959	B	T	0.100
129	N	.	-0.371	0.425	0.987	B	.	-0.300
130	P	.	-0.057	0.790	1.009	B	.	-0.150
131	V	.	-0.057	0.790	1.011	B	.	-0.150
132	S	.	-0.300	0.344	1.003	B	.	-0.150
133	N	.	-0.686	0.298	0.979	B	.	-0.600
134	F	.	-1.314	0.348	0.950	B	.	-0.600
135	I	.	-0.214	0.450	0.933	B	H	-0.300
136	S	.	-0.871	0.231	0.926	H	H	-0.600
137	F	.	-0.871	0.461	0.935	H	H	-0.600
138	E	.	0.086	1.316	0.958	H	H	0.450
139	L	.	0.829	1.417	0.987	H	H	0.750
140	E	.	0.071	1.147	1.014	H	H	0.600
141	K	.	1.029	1.325	1.037	H	H	0.900
142	T	.	1.029	2.782	1.046	H	H	0.900
143	I	.	1.029	1.325	1.041	H	H	0.900
144	K	.	0.986	0.901	1.032	H	H	0.750
145	E	.	0.929	1.082	1.010	H	H	0.900
146	L	.	0.186	1.082	0.983	H	H	0.450
147	H	.	0.229	0.401	0.960	H	H	0.300
148	E	.	-0.271	0.229	0.945	H	H	-0.300
149	I	.	-0.271	0.447	0.944	H	H	-0.300
150	V	.	0.014	0.332	0.955	H	H	0.300
151	G	.	-0.700	0.194	0.967	.	H	-0.600
152	N	.	-1.457	0.279	0.967	.	H	-0.600
153	A	.	-0.257	0.752	0.966	H	H	-0.300
154	A	.	0.843	1.269	0.968	H	H	0.750
155	A	.	1.429	1.546	0.981	H	H	0.750
156	K	.	1.114	2.398	0.999	H	H	0.750
157	D	.	0.729	1.664	0.999	H	H	0.750
158	R	.	0.386	1.222	0.986	H	B	0.450
159	Y	.	0.243	0.529	0.957	B	B	0.300
160	I	.	-0.257	0.314	0.928	B	B	-0.300
161	V	.	-1.357	0.119	0.920	B	B	-0.600
162	F	.	-1.943	0.075	0.920	B	B	-0.600
163	G	.	-2.729	0.079	0.931	B	B	-0.600
164	V	.	-1.986	0.155	0.951	B	B	-0.600
165	G	.	-0.886	0.309	0.967	.	.	-0.600
166	V	.	-1.029	0.258	0.985	h	.	-0.600
167	T	.	-1.729	0.243	0.991	h	B	-0.600
168	Q	.	-0.671	0.335	0.983	h	B	-0.600
169	L	.	-0.671	0.446	0.966	h	B	-0.600
170	I	.	-0.614	0.255	0.946	h	B	-0.600
171	H	.	-1.314	0.109	0.936	h	B	-0.600
172	G	.	-2.457	0.093	0.928	.	B	-0.600
173	L	.	-1.800	0.177	0.927	B	B	-0.600
174	V	.	-1.700	0.108	0.926	B	B	-0.600
175	I	.	-2.043	0.146	0.931	B	B	-0.600
176	S	.	-1.871	0.273	0.950	B	B	-0.600
177	L	.	-0.829	0.592	0.975	B	.	-0.600
178	S	.	-0.500	0.835	0.997	.	.	-0.600
179	P	.	0.243	0.900	1.004	T	.	0.850
180	N	G	-0.129	1.102	1.000	T	T	0.800
181	M	.	0.514	1.187	0.994	.	T	1.150
182	T	.	0.629	1.187	0.997	.	T	1.150
183	A	.	0.043	0.396	1.004	.	.	0.450
184	T	.	-0.714	0.404	1.006	.	.	-0.450
185	P	.	-0.214	0.433	1.008	T	T	0.650
186	C	.	0.186	0.742	1.005	T	T	1.250
187	A	.	0.557	0.689	1.025	.	.	0.750
188	P	.	1.014	0.891	1.058	.	.	0.750

189	Q	.	0.186	1.234	1.080	t	.	0.800
190	S	.	-0.057	0.906	1.081	t	B	0.050
191	K	.	-0.057	0.592	1.047	H	B	-0.150
192	V	.	0.171	0.465	0.992	H	B	0.300
193	V	.	-0.586	0.351	0.949	H	B	-0.600
194	A	.	-0.471	0.271	0.922	H	B	-0.600
195	H	.	-0.843	0.572	0.915	H	B	-0.600
196	A	.	-0.057	1.209	0.922	H	B	-0.150
197	P	.	0.771	1.850	0.927	T	T	1.550
198	Y	.	0.429	1.009	0.930	T	T	1.250
199	Y	.	-0.429	0.865	0.930	B	.	-0.600
200	P	.	0.471	1.096	0.934	B	.	0.450
201	V	.	0.743	1.211	0.949	B	.	0.750
202	F	.	1.057	1.338	0.979	B	T	1.150
203	R	.	0.971	1.249	1.020	B	T	1.300
204	E	.	1.300	3.366	1.058	h	T	1.300
205	Q	.	2.086	6.090	1.076	h	T	1.300
206	T	.	2.086	2.692	1.064	h	T	1.300
207	K	.	1.943	2.596	1.042	h	T	1.300
208	Y	.	2.000	2.998	1.018	h	T	1.300
209	F	.	2.057	4.154	1.014	h	T	1.300
210	D	.	2.014	2.056	1.034	h	T	1.300
211	K	.	1.643	2.056	1.045	h	T	1.300
212	K	.	1.957	4.112	1.045	h	T	1.300
213	G	.	2.486	2.589	1.025	.	T	1.300
214	Y	.	2.543	2.589	0.999	.	T	1.150
215	E	.	2.043	1.281	0.992	h	T	1.150
216	W	.	1.986	2.082	1.000	h	T	1.150
217	K	.	1.671	1.342	1.020	h	.	0.900
218	G	.	1.229	0.783	1.028	h	.	0.750
219	N	.	1.229	1.243	1.013	h	.	0.900
220	A	.	1.286	0.974	0.988	h	.	0.600
221	A	.	0.129	0.731	0.960	h	.	0.300
222	D	.	0.571	0.731	0.945	h	T	1.000
223	Y	.	0.171	1.044	0.951	.	T	0.850
224	V	.	0.543	1.385	0.978	.	T	1.150
225	N	G	0.900	1.197	1.014	.	.	0.900
226	T	.	0.629	1.181	1.050	T	.	1.300
227	S	.	0.943	2.755	1.079	T	.	1.300
228	T	.	2.043	2.967	1.083	.	.	0.900
229	P	.	1.143	1.780	1.069	T	.	1.300
230	E	.	0.400	0.931	1.041	T	.	0.850
231	Q	.	0.786	1.118	1.002	h	B	0.900
232	F	.	0.414	0.715	0.961	h	B	0.300
233	I	.	-0.414	0.307	0.932	h	B	-0.600
234	E	.	-0.814	0.255	0.922	h	B	-0.600
235	M	.	-1.200	0.395	0.937	h	B	-0.600
236	V	.	-0.571	0.872	0.975	h	.	-0.600
237	T	.	0.571	0.810	1.023	h	.	0.750
238	S	.	0.571	1.316	1.066	h	.	0.900
239	P	.	1.071	2.741	1.087	T	.	1.300
240	N	.	2.171	3.290	1.099	T	.	1.300
241	N	.	2.129	2.429	1.093	t	.	1.100
242	P	.	1.471	1.296	1.074	T	.	1.300
243	E	.	0.700	0.664	1.056	T	.	1.150
244	G	.	0.843	0.809	1.024	t	H	0.950
245	L	.	0.800	0.712	0.993	H	H	0.600
246	L	.	1.071	0.712	0.969	H	H	0.600
247	R	.	-0.029	0.534	0.949	H	H	-0.300
248	H	.	-0.729	0.454	0.941	H	H	-0.600
249	E	.	0.371	1.101	0.943	H	H	0.450
250	V	.	0.971	0.556	0.953	H	H	0.600
251	I	.	-0.029	0.219	0.968	H	H	-0.300
252	K	.	0.071	0.253	0.995	H	T	0.700
253	G	.	-0.314	0.457	1.014	.	T	0.250
254	C	.	-0.357	0.457	1.016	.	T	0.250
255	K	.	0.471	0.358	1.015	.	T	0.850
256	S	.	0.414	0.604	0.992	.	T	0.700

257	I	.	0.086	1.115	0.960	B	B	0.450
258	Y	.	-0.157	0.414	0.938	B	B	-0.300
259	D	.	-0.529	0.484	0.912	B	B	-0.600
260	M	.	-0.457	1.082	0.890	B	B	-0.450
261	V	.	0.314	0.726	0.883	B	B	0.300
262	Y	.	0.357	0.672	0.876	B	B	0.300
263	Y	.	0.314	0.924	0.882	B	B	0.300
264	W	.	0.771	1.951	0.900	B	.	0.750
265	P	.	1.471	1.797	0.922	T	.	1.150
266	H	.	1.514	1.773	0.953	T	.	1.150
267	Y	.	0.686	1.182	0.980	.	.	0.750
268	T	.	1.114	1.529	1.009	.	.	0.900
269	P	.	1.071	1.761	1.024	.	T	1.300
270	I	.	1.171	2.247	1.028	.	T	1.300
271	K	.	0.729	1.573	1.031	.	T	1.300
272	Y	.	1.129	1.699	1.028	.	.	0.900
273	K	.	1.400	4.197	1.039	.	H	0.900
274	A	.	2.543	3.505	1.054	.	H	0.900
275	D	.	1.343	1.568	1.057	.	H	0.900
276	E	.	0.886	0.776	1.040	.	H	0.750
277	D	.	-0.214	0.633	1.004	.	H	-0.150
278	I	.	-0.357	0.328	0.955	B	H	-0.300
279	M	.	-0.757	0.274	0.918	B	H	-0.600
280	L	.	-1.529	0.162	0.908	B	H	-0.600
281	F	.	-1.914	0.310	0.916	B	H	-0.600
282	T	.	-0.714	0.627	0.946	B	H	-0.600
283	M	.	-0.257	1.191	0.975	B	H	-0.150
284	S	.	0.386	1.984	1.002	t	.	0.800
285	K	.	0.843	1.361	1.020	t	T	1.500
286	Y	.	1.200	1.871	1.022	t	T	1.500
287	T	.	1.586	1.871	1.027	t	.	1.100
288	G	.	1.529	0.926	1.032	t	.	0.950
289	H	.	1.086	0.792	1.045	t	.	0.950
290	S	.	1.543	1.075	1.067	T	.	1.300
291	G	.	1.043	0.940	1.076	T	.	1.150
292	S	.	1.043	0.684	1.065	t	.	0.950
293	R	.	0.714	0.537	1.031	t	T	1.350
294	F	.	0.343	0.548	0.986	t	T	0.900
295	G	.	-0.257	0.337	0.948	.	T	0.100
296	W	.	-1.014	0.121	0.921	H	.	-0.600
297	A	.	-1.100	0.279	0.920	H	.	-0.600
298	L	.	-0.200	0.470	0.937	H	.	-0.300
299	I	.	0.243	0.774	0.965	H	.	0.300
300	K	.	0.214	1.106	1.005	H	H	0.600
301	D	.	-0.129	0.996	1.025	H	H	-0.150
302	E	.	0.600	2.225	1.029	H	H	0.900
303	T	.	1.743	1.789	1.018	B	H	0.900
304	V	.	1.743	2.143	0.998	B	H	0.750
305	Y	.	0.700	1.020	0.990	B	H	0.750
306	N	.	-0.343	0.583	0.991	B	H	-0.300
307	K	.	0.057	1.263	0.992	B	H	0.450
308	L	.	0.843	1.263	0.982	B	H	0.750
309	L	.	0.386	0.777	0.963	B	H	0.300
310	N	.	-0.014	0.561	0.948	B	H	-0.300
311	Y	.	-0.014	1.361	0.948	B	H	-0.150
312	M	.	1.029	2.653	0.970	B	.	0.750
313	T	.	1.671	2.381	1.010	B	.	0.900
314	K	.	1.671	2.632	1.048	t	.	1.100
315	N	.	1.543	2.632	1.079	t	.	1.100
316	T	.	1.914	2.632	1.095	T	.	1.300
317	E	.	1.929	1.763	1.104	T	.	1.300
318	G	.	2.014	2.148	1.107	t	.	1.100
319	T	.	2.014	2.577	1.104	t	.	1.100
320	S	.	2.014	2.148	1.103	t	.	1.100
321	R	.	2.014	3.759	1.088	h	H	0.900
322	E	.	1.414	2.148	1.071	h	H	0.900
323	T	.	1.957	3.139	1.054	h	H	0.900
324	Q	.	1.957	2.148	1.035	h	H	0.900

325	L	.	0.771	1.023	1.021	h	H	0.900
326	R	.	0.829	1.417	1.017	h	H	0.900
327	S	.	0.086	0.574	1.009	h	H	0.450
328	L	.	-0.957	0.574	0.997	h	H	-0.600
329	K	.	0.143	0.586	0.996	h	H	0.300
330	I	.	0.000	0.757	0.988	h	H	-0.300
331	L	.	-0.714	0.681	0.990	h	H	-0.600
332	K	.	-0.814	0.239	0.993	h	H	-0.600
333	E	.	-1.629	0.344	0.976	h	H	-0.600
334	V	.	-1.257	0.413	0.947	B	H	-0.600
335	I	.	-1.314	0.153	0.915	B	H	-0.600
336	A	.	-1.314	0.177	0.901	B	H	-0.600
337	M	.	-1.714	0.344	0.911	B	H	-0.600
338	V	.	-0.614	0.850	0.958	B	H	-0.600
339	K	.	0.586	1.683	1.016	B	H	0.900
340	T	.	0.900	1.683	1.064	B	H	0.900
341	Q	.	1.271	3.273	1.099	t	H	1.100
342	K	.	1.600	1.619	1.098	t	.	1.100
343	G	.	1.686	2.198	1.083	t	.	1.100
344	T	.	2.086	2.119	1.060	.	B	0.900
345	M	.	1.043	0.874	1.034	.	B	0.750
346	R	.	0.986	1.420	1.025	.	B	0.900
347	D	.	1.029	1.420	1.022	.	B	0.900
348	L	.	0.529	1.243	1.012	.	.	0.900
349	N	.	0.857	0.628	1.004	.	.	0.750
350	T	.	-0.186	0.326	0.988	T	.	0.100
351	F	.	-0.186	0.684	0.969	T	.	0.100
352	G	.	0.914	0.850	0.970	.	.	0.600
353	F	.	-0.129	0.486	0.976	H	.	-0.300
354	Q	.	0.414	1.099	0.993	H	.	0.450
355	K	.	1.314	1.923	1.015	H	.	0.900
356	L	.	1.900	4.350	1.025	H	.	0.900
357	R	.	2.429	2.641	1.029	H	.	0.900
358	E	.	1.329	0.980	1.020	H	T	1.150
359	R	.	1.271	1.911	0.997	H	T	1.150
360	W	.	1.171	0.684	0.976	B	T	1.000
361	V	.	0.629	0.570	0.966	B	.	0.600
362	N	G	0.243	0.390	0.970	B	.	0.300
363	I	.	-0.943	0.306	0.986	B	.	-0.600
364	T	.	-1.614	0.340	1.002	B	.	-0.450
365	S	.	-0.514	0.353	1.011	B	.	-0.450
366	L	.	-0.457	1.007	1.015	B	.	-0.300
367	L	.	0.300	0.935	1.031	B	.	0.450
368	D	.	0.700	1.165	1.054	.	T	1.300
369	K	.	1.229	2.768	1.070	T	T	1.700
370	S	.	1.371	2.906	1.074	T	T	1.700
371	D	.	2.029	2.332	1.049	t	T	1.500
372	R	.	1.714	1.827	1.015	b	T	1.300
373	F	.	1.657	2.361	0.985	b	T	1.150
374	S	.	2.100	2.827	0.975	b	T	1.150
375	Y	.	1.057	1.191	0.982	b	T	1.150
376	Q	.	0.643	2.126	1.002	b	T	1.300
377	K	.	1.543	2.747	1.031	b	.	0.900
378	L	.	1.543	2.350	1.055	b	.	0.900
379	P	.	1.857	2.350	1.081	.	T	1.300
380	Q	.	1.543	1.841	1.090	t	T	1.500
381	S	.	0.629	1.197	1.072	t	T	1.500
382	E	.	1.671	1.245	1.034	t	T	1.500
383	Y	.	1.629	1.126	0.985	t	T	1.350
384	C	.	0.729	0.728	0.944	t	T	1.200
385	N	.	1.257	0.823	0.931	t	T	1.200
386	Y	.	1.400	1.029	0.938	B	T	1.150
387	F	.	0.943	1.899	0.954	B	T	1.150
388	R	.	1.943	2.313	0.978	B	T	1.150
389	R	.	1.671	2.282	0.994	B	T	1.150
390	M	.	1.714	4.076	1.006	B	T	1.300
391	R	.	2.229	2.789	1.030	.	.	0.900
392	P	.	1.814	2.201	1.049	t	.	1.100

393	P	.	1.286	2.981	1.066	T	.	1.300
394	S	.	1.743	2.385	1.073	T	.	1.300
395	P	.	0.843	1.558	1.049	T	T	1.700
396	S	.	0.743	1.060	1.008	T	T	1.700
397	Y	.	-0.086	0.587	0.960	B	T	0.100
398	A	.	0.357	0.759	0.927	B	T	0.700
399	W	.	-0.229	0.304	0.916	B	.	-0.300
400	V	.	0.157	0.336	0.928	B	H	0.300
401	K	.	0.100	0.349	0.949	B	H	0.300
402	C	.	0.857	0.575	0.960	B	H	0.600
403	E	.	1.229	1.342	0.975	H	H	0.750
404	W	.	2.329	1.121	0.988	H	H	0.750
405	E	.	2.329	4.181	1.012	H	H	0.900
406	E	.	3.186	4.032	1.040	H	H	0.900
407	D	.	2.329	2.056	1.047	H	T	1.300
408	K	.	2.386	1.860	1.041	H	T	1.300
409	D	.	2.386	1.860	1.020	H	T	1.300
410	C	.	1.986	1.607	0.993	B	T	1.150
411	Y	.	1.086	0.696	0.984	B	T	1.000
412	Q	.	1.029	0.722	0.987	B	T	1.000
413	T	.	1.029	2.165	0.999	B	T	1.150
414	F	.	1.443	1.367	1.020	B	T	1.300
415	Q	.	1.900	1.546	1.043	T	T	1.700
416	N	.	0.757	0.751	1.059	T	T	1.550
417	G	.	1.157	1.395	1.062	t	T	1.500
418	R	.	1.657	1.162	1.056	t	T	1.500
419	I	.	1.657	1.252	1.051	b	.	0.900
420	N	.	1.657	2.034	1.056	b	.	0.900
421	T	.	1.657	1.028	1.076	b	.	0.900
422	Q	.	1.514	2.539	1.101	b	.	0.900
423	N	.	2.214	1.563	1.111	b	.	0.900
424	G	.	1.314	0.938	1.101	T	.	1.150
425	E	.	1.714	0.938	1.070	T	.	1.150
426	G	.	0.957	0.589	1.034	t	.	0.950
427	F	.	0.514	0.589	1.006	.	.	0.750
428	E	.	0.571	0.456	1.005	.	.	0.750
429	A	.	0.714	0.902	1.022	t	.	0.950
430	G	.	0.843	1.632	1.036	t	T	1.500
431	S	.	0.643	0.700	1.043	T	T	1.550
432	R	.	0.786	1.356	1.018	T	T	1.700
433	Y	.	0.500	1.130	0.983	B	T	1.150
434	V	.	0.557	1.130	0.955	B	B	0.750
435	R	.	-0.100	0.476	0.936	B	B	-0.300
436	L	.	-1.386	0.213	0.934	B	B	-0.600
437	S	.	-1.014	0.574	0.942	B	B	-0.600
438	L	.	-0.314	0.423	0.960	B	B	-0.300
439	I	.	-0.400	1.025	0.984	B	B	-0.450
440	K	.	0.643	1.277	1.018	.	B	0.900
441	T	.	1.029	2.587	1.045	.	T	1.300
442	K	.	1.171	3.195	1.053	t	T	1.500
443	D	.	2.314	2.668	1.053	t	T	1.500
444	D	.	2.257	3.202	1.044	t	H	1.100
445	F	.	1.614	1.320	1.027	t	H	1.100
446	D	.	0.786	0.782	1.010	t	H	0.950
447	Q	.	0.471	0.734	0.979	B	H	0.300
448	L	.	0.157	1.328	0.935	B	H	0.450
449	M	.	0.014	0.656	0.899	B	H	0.300
450	Y	.	0.071	0.758	0.889	B	H	0.300
451	Y	.	0.071	1.477	0.905	B	H	0.450
452	L	.	0.343	1.477	0.935	B	H	0.450
453	K	.	0.014	0.700	0.962	.	H	0.300
454	N	.	0.329	0.773	0.963	.	H	0.300
455	M	.	-0.114	0.947	0.955	H	H	-0.300
456	V	.	0.986	0.947	0.959	H	H	0.600
457	E	.	1.071	1.154	0.973	H	H	0.750
458	A	.	1.129	2.332	1.006	H	H	0.900
459	K	.	1.500	4.534	1.039	H	H	0.900
460	R	.	2.329	4.048	1.058	H	H	0.900

461	K	.	1.286	3.305	1.065	H	H	0.900
462	T	.	0.900	1.158	1.050	H	H	0.900
463	P	.	0.900	1.183	1.033	.	H	0.900
464	L	.	0.757	1.024	1.013	B	B	0.900
465	I	.	-0.343	0.585	0.998	B	B	-0.300
466	K	.	-0.329	0.507	1.005	B	B	-0.150
467	Q	.	-0.057	0.989	1.014	B	B	-0.150
468	L	.	0.986	2.356	1.026	B	B	0.900
469	S	.	2.129	2.041	1.042	t	.	1.100
470	N	.	0.929	0.826	1.046	T	T	1.550
471	D	.	0.543	1.342	1.043	T	T	1.700
472	Q	.	1.729	1.962	1.040	.	T	1.300
473	I	.	2.257	2.389	1.033	.	T	1.300
474	S	.	1.986	2.212	1.035	.	T	1.300
475	R	.	1.086	1.106	1.033	t	T	1.500
476	R	.	-0.057	1.106	1.000	t	.	0.200
477	P	.	0.683	1.055	1.000	.	T	1.300
478	F	.	0.660	0.689	1.000	.	T	1.150
479	I	.	-0.300	0.449	1.000	.	.	-0.150